

Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft

Band 54

Semantische Relationen in Dokumentationssprachen –
vom Thesaurus zum semantischen Netz

Felix Boteram

November 2008

Fachhochschule Köln
Fakultät für Informations- und Kommunikationswissenschaften
Institut für Informationswissenschaft

Boteram, Felix
Semantische Relationen in Dokumentationssprachen – vom Thesaurus zum
semantischen Netz
Köln : Fachhochschule Köln,
Fakultät für Informations- und Kommunikationswissenschaften,
Institut für Informationswissenschaft, 2008
(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 54)

ISSN (elektronische Version) 1434-1115

Die Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft berichten
über aktuelle Forschungsergebnisse des Instituts Informationswissenschaft der Fach-
hochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozentinnen und Dozen-
ten als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaft-
lichen Qualität der Veröffentlichungen liegt bei der Schriftleitung.
Jeder Band erscheint in elektronischer Version
(im WWW: <http://www.fbi.fh-koeln.de/institut/papers/arbeitspapiere.php>).

Fachhochschule Köln
Fakultät für Informations- und Kommunikationswissenschaften
Institut für Informationswissenschaft
Claudiusstr.1 D-50678 Köln
Tel.: 0221/8275-3376, Fax: 0221/3318583
E-Mail: schriftenreihe@fbi.fh-koeln.de

Schriftleitung: Prof. Dr. Konrad Scherfer
© FH Köln 2008

ABSTRACT

Moderne Verfahren des Information Retrieval verlangen nach aussagekräftigen und detailliert relationierten Dokumentationssprachen. Der selektive Transfer einzelner Modellierungsstrategien aus dem Bereich semantischer Technologien für die Gestaltung und Relationierung bestehender Dokumentationssprachen wird diskutiert. Am Beispiel des Gegenstandsbereichs „Theater“ der Schlagwortnormdatei wird ein hierarchisch strukturiertes Relationeninventar definiert, welches sowohl hinreichend allgemeine als auch zahlreiche spezifische Relationstypen enthält, welche eine detaillierte und damit funktionale Relationierung des Vokabulars ermöglichen. Die Relationierung des Gegenstandsbereichs wird als Ontologie im OWL-Format modelliert. Im Gegensatz zu anderen Ansätzen und Überlegungen zur Schaffung von Relationeninventaren entwickelt der vorgestellte Vorschlag das Relationeninventar aus der Begriffsmenge eines bestehenden Gegenstandsbereichs heraus. Das entwickelte Inventar wird als eine hierarchisch strukturierte Taxonomie gestaltet, was einen Zugewinn an Übersichtlichkeit und Funktionalität bringt.

INHALTSVERZEICHNIS

1	EINLEITUNG	6
2	DOKUMENTATIONSSPRACHEN UND INFORMATION RETRIEVAL	8
2.1	Externalisierung von Wissen	8
2.2	Gestaltung von Dokumentationssprachen	10
2.3	Erfordernisse des Information Retrieval	12
2.4	Strukturen und Typen von Dokumentationssprachen	16
2.4.1	Klassifikationen	18
2.4.2	Thesauri	20
2.4.3	Hybridformen	20
2.4.4	Semantische Netze	21
3	THESAURI	22
3.1	Herkömmliche Relationstypen	23
3.2	Nicht-konventionelle Relationstypen	26
4	SEMANTISCHE NETZE IM INFORMATION RETRIEVAL	29
4.1	Gestaltungsprinzipien semantischer Netze	32
4.1.1	Klassenbildung	33
4.1.2	Typisierung von Relationen	33
4.2	Erweiterung von Retrievalfunktionalitäten durch semantische Technologien	35
4.2.1	Orientierungsfunktion („Query-clarification“)	35
4.2.2	Modifikation („Query-modification“)	36
4.2.3	Expansion („Query-expansion“)	37
5	TYPLOGIE SEMANTISCHER NETZE	39
6	EXEMPLARISCHE RELATIONIERUNG EINES TEILGEBIETS DER SWD	41
6.1	Material / Ausgangslage	44
6.2	Werkzeuge / Software	48
6.3	Technisches Vorgehen	50
6.4	Bisherige Entwicklungen	52
6.5	Strukturierung des Gegenstandsbereichs	55
6.5.1	Begriffsdefinition durch Klassenbildung	56
6.5.2	Semantische Relationen	57
6.6	Definition und Strukturierung eines Relationeninventars	57
6.6.1	Reduktions- und Abstraktionsprozesse	59
6.6.2	Typisierung der Relationsarten	61
6.7	Darstellung und Anwendung der neuen Strukturen unter Protégé	73
6.8	Regeln für die Anwendung des Relationeninventars	77
7	SCHNITTSTELLEN UND PRODUKTE	80
7.1	Schnittstellen für Visualisierungs- und Navigationskomponenten	80
7.2	Darstellung der Retrievalmöglichkeiten für den Benutzer	81

7.3	Weiterführende Überlegungen / Ausblick.....	81
8	SCHLUSS	82
	LITERATURVERZEICHNIS	83
	Internetquellen.....	86
	Referenzwerke	86
	Unveröffentlicht	86
	ANHANG	87

ABBILDUNGSVERZEICHNIS

Abbildung 1: Zusammenhang zwischen Information und Wissen	9
Abbildung 2: „semantische Treppe“	18
Abbildung 3: Topterm „Theater“ in Listendarstellung (links).....	47
Abbildung 4: Topterm „Theater“ in Baumstruktur (rechts).....	47
Abbildung 5: unspezifische Relationierung mittels eines Relationstyps	74
Abbildung 6: differenzierte Relationierung des Terms „Theater“ durch neue Relationstypen	75
Abbildung 7: spezifische Auswahl eines Relationstyps.....	76
Abbildung 8: spezifische Auswahl eines Relationstyps.....	76
Abbildung 9: unmittelbares semantisches Umfeld des Begriffs „Bühnenbeleuchtung“	77
Abbildung 10: Selektive Relationierung des Terms "Beruf"	90
Abbildung 11: Selektive Relationierung des Terms "Beruf"	90
Abbildung 12: Hierarchieleiter zum Term "Beruf" mit Relationstypen unterschiedlicher Spezifität	91
Abbildung 13: Hierarchieleiter zum Term "Beruf" Hierarchieleiter zum Term „Beruf“	91
Abbildung 14: Semantisches Umfeld des Terms "Beleuchter"	92
Abbildung 15: Relation "Beteiligte " zum Term "Commedia dell'arte" .	92
Abbildung 16: Semantisches Umfeld des Terms "Oper"	93

ABKÜRZUNGSVERZEICHNIS

ALA	American Library Association
CRG	Classification Research Group
DDC	Dewey Decimal Classification
MACS	Multilingual Access to Subjects
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RSWK	Regeln für den Schlagwortkatalog
SKOS	Simple Knowledge Organization System
SWD	Schlagwortnormdatei
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1 EINLEITUNG

Ausgehend von der Erkenntnis, dass sich Methoden und Werkzeuge der inhaltlichen Erschließung an den Besonderheiten und Erfordernissen des Information Retrieval in elektronischen Umgebungen orientieren müssen, sollen im Rahmen dieser Arbeit die Möglichkeiten zur Erweiterung der Retrievalfunktionalitäten durch die Integration umfangreicher und differenziert relationierter Dokumentationssprachen in den Erschließungsprozess untersucht werden. Zu Beginn sollen verschiedene Typen von Dokumentationssprachen und ihre Strukturen und Eigenschaften dargestellt werden. Es soll untersucht werden, welche dieser Eigenschaften die gewünschten Retrievalfunktionalitäten unterstützen. Dabei soll gezeigt werden, wie Dokumentationssprachen und Erschließungsvokabulare durch Anreicherung um zusätzliche Strukturmerkmale und Relationen erweitert werden können, um auf diese Weise die Realisation verschiedener zusätzlicher Suchmöglichkeiten für den Benutzer zu ermöglichen.

Insbesondere die Kombination verschiedener Methoden sowie die Erweiterung bereits differenziert entwickelter Thesauri zu semantischen Netzen soll genauer erörtert werden. Dabei soll näher auf die Struktur und Funktionsweise von semantischen Netzen sowie deren Besonderheiten in Bezug auf das Information Retrieval eingegangen werden. Hierbei ist von besonderem Interesse, welche Retrievalfunktionalitäten den Benutzern angeboten werden können: Neben Suchfunktionalitäten wie der Modifikation oder der gezielten und kontrollierten Expansion von Suchanfragen soll auch die Orientierungsfunktion von Dokumentationssprachen bei der Klärung unspezifischer Informationsbedarfe betont werden.

Hierzu soll eine Typisierung und systematische Ordnung aller vorhandenen Relationen sowie die detaillierte Darstellung ihrer logischen Eigenschaften verwendet werden.

Um die Vielfalt aller möglichen neuen Relationen die in einem semantischen Netz definiert werden können überschaubar zu halten und die Komplexität dieser Systeme für den Benutzer handhabbar zu gestalten, ist eine Beschränkung auf eine sinnvolle Anzahl von Relationen notwendig. Dies ist besonders wichtig, da die hier behandelten semantischen Netze zur Unterstützung von Benutzern über grafische Benutzeroberflächen und nicht als Grundlage für die Arbeit „intelligenter Agenten“ konzipiert werden.

Welche Überlegungen einem solchen Reduktions- und Auswahlprozess zugrunde liegen und welche Eigenschaften von Relationen am sinnvollsten sind, soll dabei ausführlich dargestellt werden.

Nach der überblicksartigen Darstellung verschiedener Arten von Dokumentationssprachen im zweiten Kapitel sollen im darauf folgenden dritten Kapitel Thesauri sowie deren Relationenstruktur untersucht werden. Im vierten Kapitel sollen Eigenschaften und Funktionsweisen semantischer

Netze erörtert, und die Übertragbarkeit einzelner semantischer Technologien und Gestaltungsmöglichkeiten untersucht werden. Anhand dieser Überlegungen soll im dann im fünften Kapitel beispielhaft ein Relationeninventar aus einem konkreten Gegenstandsbereich der Schlagwortnormdatei herausmodelliert und entwickelt werden.

Produkte und Funktionalitäten, die auf solchen semantischen Strukturen aufsetzen können und weitergehenden Forschungs- und Entwicklungsfragen, welche die angestellten Überlegungen aufgreifen können, werden in einem Ausblick angesprochen.

2 DOKUMENTATIONSSPRACHEN UND INFORMATION RETRIEVAL

Die Erschließung von Inhalten bedient sich in den allermeisten Fällen des Hilfsmittels der Sprache. Neben natürlichsprachlichen Ansätzen wie dem Verfahren des „Abstractings“ bei dem Inhalte in kurze prägnante Texte in natürlicher Sprache zusammengefasst werden, stehen zahlreiche Verfahren die sich verschiedener Formen konstruierter Sprachen bedienen.

In diesem Kapitel sollen Eigenschaften und Funktionen von Dokumentationssprachen sowie deren Rolle im Prozess des Information Retrievals dargestellt werden.

Hierbei sollen einige Besonderheiten von Dokumentationssprachen durch einen Vergleich mit der natürlichen Sprache verdeutlicht werden.

2.1 Externalisierung von Wissen

Um Prinzipien der Wissensorganisation und Wissensrepräsentation sowie Verfahren der Erschließung von Wissen verstehen und diskutieren zu können, scheint eine begriffliche Klärung der Konzepte von „Wissen“ sowie von „Information“ Voraussetzung. Daher sollen im Folgenden verschiedene Definitionsansätze sowie grundlegende Unterschiede dieser beiden zentralen Begriffe herausgearbeitet werden, sowie die Zusammenhänge und deren Relevanz für den Bereich der Wissensrepräsentation aufgezeigt werden.

Scheint eine präzise Definition des Begriffs des Wissens kaum möglich, so lassen sich doch einige wichtige Eigenschaften herausarbeiten, vor allem, wenn er in Zusammenhang mit und im Gegensatz zu dem Begriff der Information gesehen wird.

Wissen entsteht erst durch die Verarbeitung und Interpretation von Information durch den einzelnen Rezipienten. Es ist daher abhängig von den kognitiven Strukturen des Einzelnen und dem jeweiligen Kontext in dem dieses Wissen erworben oder konstruiert wird.

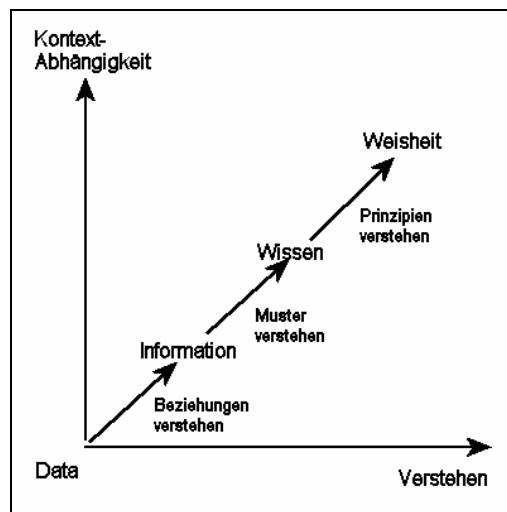


Abbildung 1: Zusammenhang zwischen Information und Wissen¹

Voraussetzung für eine Speicherung und Vermittlung solchen Wissens ist die Externalisierung dieser Strukturen, in eine intersubjektiv nachvollziehbare und somit vermittelbare und nachprüfbare Darstellung der Bedeutungsstrukturen und der Zusammenhängen und Bezügen zwischen einzelnen Elementen des Wissens. Dabei ist die Vorstellung der Möglichkeit einer objektiven und von kognitiven Strukturen Einzelner unabhängigen Externalisierung von Wissen problematisch.² Die Aufbereitung, Darstellung, und vor allem die Rezeption von Information und damit auch von Wissen ist immer ein Vermittlungsprozess, der durch die kognitiven Strukturen vor allem des Rezipienten sowie dessen Fähigkeit, neue Elemente in vorhandene Strukturen einzupassen beeinflusst wird und von diesen abhängt. Wichtig ist in diesem Zusammenhang die „Offenlegung des erkenntnistheoretischen Modells“³ sowie die gründliche Reflexion der zur Kommunikation dieses Wissens gebrauchten begrifflichen Strukturen als Teil der verwendeten Sprache sowie deren transparente Darstellung für alle Beteiligten. Die Gestaltung dieser Sprache ist für die Möglichkeiten der Externalisierung und Kommunikation von entscheidender Bedeutung.

¹ Abbildung von <http://www.brainworker.ch/km> übernommen.

² Zur Problematik verschiedener Auffassungen zur Möglichkeit eines objektivierten Wissens sowie dessen Vermittlung s. Gödert, Winfried. Externalisierung von Wissen: Eine informationstheoretische Betrachtung aus konstruktivistischer Sicht. In: Meder, Norbert; Peter Jaenecke, Winfried Gödert (Hrsg.): Konstruktion und Retrieval von Wissen. 3. Tagung der Deutschen ISKO Sektion einschließlich der Vorträge des Workshops „Thesauri als terminologische Lexika.“ Weilburg, 27.-29. 10. 1993. Frankfurt a.M.: Indeks-Verlag 1995. (Fortschritte in der Wissensorganisation 3).

³ Gödert, Winfried. Externalisierung von Wissen: Eine informationstheoretische Betrachtung aus konstruktivistischer Sicht. In: Meder, Norbert; Peter Jaenecke, Winfried Gödert (Hrsg.): Konstruktion und Retrieval von Wissen. 3. Tagung der Deutschen ISKO Sektion einschließlich der Vorträge des Workshops „Thesauri als terminologische Lexika.“ Weilburg, 27.-29. 10. 1993. Frankfurt a.M.: Indeks-Verlag 1995. (Fortschritte in der Wissensorganisation 3) S. 13

Für Dokumentationssprachen, welche nicht durch natürliche Bildungs- und Entwicklungsprozesse geformt werden, sondern bewusst und planvoll konstruiert und angewendet werden müssen, bedeutet dies die Notwendigkeit der sorgfältigen Konstruktion und transparenten Darstellung der von ihr verwendeten Vokabularen. Soll diese Externalisierung und Vermittlung zudem noch durch eine maschinelle Verarbeitung in elektronischen Umgebungen wie Information Retrieval Systemen unterstützt werden, ist zudem eine formale Strenge sowie logische Validität bei der Gestaltung dieser Dokumentationssprachen, vor allem ihrer Vokabularen notwendig.

2.2 Gestaltung von Dokumentationssprachen

Dokumentationssprachen sind künstliche Sprachen, welche zur Erschließung, Ordnung und zum Retrieval d.h. dem gezielten Wiederauffinden von Dokumenteninhalten oder deren Repräsentanten innerhalb eines Dokumentationssystems verwendet werden.

Wie natürliche Sprachen dienen sie der Kommunikation and damit der Vermittlung von Bedeutung. Im Falle von Dokumentationssprachen findet diese Kommunikation innerhalb eines klar umrissenen Systems statt, mit dem Ziel, für eine bestimmte Fragestellung relevante Dokumente oder die darin enthaltenen Informationen (wieder) zu finden. Sie sind als eigene künstlich geschaffene, konstruierte und hoch entwickelte Sprachen zu begreifen,⁴ die für das Information Retrieval sowie die Wissensrepräsentation und -organisation Funktionsgrundlage und nicht nur Hilfsmittel sind.

Im Folgenden sollen verschiedene Definitionen und Sichtweisen dargestellt werden und eine begriffliche Klärung darüber stattfinden, was im Rahmen dieser Arbeit unter Dokumentationssprachen verstanden wird und welche Eigenschaften und Funktionsweisen von Dokumentationssprachen dabei von besonderem Interesse sind. Sowohl die Bedeutung als auch die Komplexität des Gegenstands der Dokumentationssprachen im Bereich der Wissensorganisation und des Information Retrieval zeigt sich in einer Vielzahl unterschiedlicher Definitionen und Begrifflichkeiten.⁵ Während bei Foscett⁶ Eigenschaften und Bestandteile von Dokumentationssprachen behandelt werden, betonen andere die Funktion von Dokumentationssprachen. So lenkt Hutchins die Aufmerksamkeit auf die kommunikative Funktion

⁴ Vgl. Gödert, Winfried. Verbale Inhaltserschließung : Ein Übersichtsartikel als kommentierter Literaturbericht. In: Mitteilungen für Bibliotheken des Landes Nordrhein-Westfalen, 41 (1991) 1, S. 1-21 hier besonders S. 3

⁵ Vgl. Bertram, Jutta. Einführung in die inhaltliche Erschließung. Grundlagen, Methoden, Instrumente. Würzburg: Ergon.-Verl. 2005., S. 127 siehe hier auch Fußnoten S. 127 hier insbesondere Gödert, Winfried. Verbale Inhaltserschließung : Ein Übersichtsartikel als kommentierter Literaturbericht. In: Mitteilungen für Bibliotheken des Landes Nordrhein-Westfalen, 41 (1991) 1, S. 1-21 und Soergel, Dagobert, Klassifikationssysteme und Thesauri. Eine Anleitung zur Herstellung von Klassifikationssystemen und thesauri im Bereich der Dokumentation. Frankfurt am Main: Deutsche Gesellschaft für Dokumentation 1969. S. 40, Abschnitt B 3.0

⁶ Vgl. Foscett, A. C. The subject approach to Information. 5th edition, London: Library Association Publishing, 1996, S. 113 ff.

von Dokumentationssprachen während Soergel⁷ den Aspekt der „Wiederauffindbarkeit“ der erschlossenen Dokumente betont. Die Betrachtung von Dokumentationssprachen unter dem Gesichtspunkt ihrer Funktion für das Wiederauffinden von Informationen innerhalb eines Dokumentationssystems ist ein Hauptaspekt dieser Arbeit und soll später genauer betrachtet werden.

Auch wenn der Begriff der Dokumentationssprache vorrangig Methoden zur verbalen Erschließung zu implizieren scheint, wird unter Dokumentationssprachen die Gesamtheit aller Systeme und Hilfsmittel verstanden, die zur Erschließung, Ordnung und Speicherung von Informationen genutzt werden, und somit als „dokumentarische Ordnungssysteme“⁸ fungieren können. So sollen hier neben Erschließungsvokabularen wie reinen Schlagwortlisten und Thesauren auch Klassifikationssysteme und Ontologien als Grundlage für semantische Netze als Dokumentationssprachen verstanden und behandelt werden.

Entscheidend ist das Vorhandensein einer künstlichen, konstruierten Struktur, welche alle zum Zweck der Erschließung gebrauchten Elemente (Terme oder Klassen) einer begrifflichen und terminologischen Normierung unterzieht und ordnet. Verfahren die sich *keines* kontrollierten Vokabulars oder klar definierter Strukturen bedienen, wie Verfahren deren Termmenge sich aus Stichwörtern und somit nicht aus Elementen eines kontrollierten Vokabulars zusammensetzt, sollen im Rahmen dieser Arbeit nicht behandelt werden, auch wenn sie teilweise als Dokumentationssprachen bezeichnet oder mit diesen in Verbindung gebracht werden.⁹

Sowohl für klassifikatorische als auch für verbale Verfahren der inhaltlichen Erschließung ist die verwendete Dokumentationssprache von zentraler Bedeutung. Bei Verfahren der verbalen Erschließung bietet die Dokumentationssprache die terminologische Grundlage für Indexierung und Retrieval. Die terminologische Kontrolle durch ein begrenztes, strukturiertes Inventar an Begriffen ermöglicht es, sicherzustellen, dass auf beiden Seiten des Retrievalprozesses das identische Vokabular verwendet wird, und somit eine eindeutige, präzise und aussagekräftige Darstellung sowohl der Dokumentinhalte als auch der Anfragen von Seiten der Nutzer möglich ist. Dies ist Voraussetzung dafür, dass Indexterme mit Suchtermen oder bei syntaktischen Verfahren komplex strukturierte Indexate mit Suchanfragen abgeglichen werden können und die dabei gefundenen Übereinstimmungen zur Darstellung relevanter Treffer genutzt werden können. Dokumentationsspra-

⁷ Soergel, Dagobert, Klassifikationssysteme und Thesauri. Eine Anleitung zur Herstellung von Klassifikationssystemen und Thesauri im Bereich der Dokumentation. Frankfurt am Main: Deutsche Gesellschaft für Dokumentation 1969. S. 40, Abschnitt B 3.0

⁸Budin, Gerhard. Wissensorganisation und Terminologie. Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. Tübingen: Narr, 1996. (Forum für Fachsprachenforschung 28., S. 64

⁹ vgl. Grundlagen der praktischen Information und Dokumentation. Kuhlen, Rainer; Thomas Seeger; Dietmar Strauch (Hrsg.). 5., völlig neu gefasste Ausgabe. Bd. 2 Glossar München: Saur, 2004.S.29 Lemma „Dokumentationssprache“

chen haben somit eine kommunikative Aufgabe und funktionieren als Medium¹⁰. Das bedeutet auch, dass sie immer in ihrem gesamten kommunikativen Kontext und in ihrer Funktion im Information Retrieval gesehen werden müssen. Im Gegensatz zu natürlichen Sprachen ist dieser Kontext mit seinen strukturellen und technischen Anforderungen klar eingegrenzt und im Idealfall schon vor der Schaffung einer Dokumentationssprache genau vorgegeben. Dies ist notwendig, da die Eigenschaften und Strukturen der Dokumentationssprache genau an die Funktionalitäten des Retrievalsystems angepasst werden müssen, um ein differenziertes und effektives Retrieval zu ermöglichen.

Die durch ein durchdachtes und differenziert gestaltetes Begriffssystem geleistete terminologische Kontrolle ist für begriffsorientierte Verfahren, also Methoden, bei denen die Erschließung und Suche auf die Begriffsebene und nicht auf die sprachliche Oberflächenstruktur von Zeichenketten abzielt eine unabdingbare Voraussetzung für das Funktionieren eines Dokumentations- bzw. Information Retrieval Systems. Daher ist die Dokumentationssprache ein integraler Bestandteil eines Information Retrieval Systems und für eine funktionale Erschließung und eine erfolgreiche Recherche von zentraler Bedeutung. Sowohl Konzeption, Erstellung und Pflege als auch die Anwendung der Dokumentationssprache im Rahmen der Erschließung und der Recherche müssen diesem Zweck gerecht werden und dürfen nicht isoliert betrachtet werden. „Die Funktion eines Thesaurus muss immer im Zusammenhang mit dem gesamten Dokumentationssystem in dem er verwendet werden soll gesehen werden.“¹¹ Die Berücksichtigung dieses Systembezugs ist besonders bei hoch entwickelten und komplexen Verfahren wichtig. Dokumentationssprachen ausschließlich im Kontext des Indexierens zu sehen und das Augenmerk vorrangig auf das einzelne Indexat zu richten, ist daher problematisch; der Kontext des Retrievals und der Gebrauch der Dokumentationssprache bei der Recherche in den mit ihr erschlossenen Beständen durch den Benutzer sind immer mit zu berücksichtigen.

In welcher Form die oben genannten Anforderungen in verschiedenen Arten von Dokumentationssprachen im Einzelnen realisiert werden soll in Abschnitt 2.4 genauer dargestellt werden.

2.3 Erfordernisse des Information Retrieval

Aufgrund der Bedeutung des beschriebenen Systembezugs sowie der Funktion der Dokumentationssprache als Retrievalsprache, sind die Erfordernisse

¹⁰ vgl. Hutchins, W. John, Languages of indexing and classification. A linguistic study of structures and functions. Stevenage: Peregrinus 1975 S. 11, Absatz 2.8.8. Der von Hutchins verwendete Begriff des „Kanals“ ist angesichts der Unzulänglichkeit der „Conduit-Metapher“ allerdings problematisch.

¹¹ Soergel, Dagobert, Klassifikationssysteme und Thesauri. Eine Anleitung zur Herstellung von Klassifikationssystemen und thesauri im Bereich der Dokumentation. Frankfurt am Main: Deutsche Gesellschaft für Dokumentation 1969. S. 30, A 1.5 Auch wenn die Begriffe Thesaurus und Klassifikationssystem nicht wie von Soergel beobachtet synonym verwendet werden sollen, lässt sich diese Aussage auf alle Arten von Dokumentationssprachen übertragen.

des Information Retrievals von zentraler Bedeutung. Die Gestalt der Dokumentationssprache muss an Funktionserfordernisse und Systemparameter des Retrievals angepasst und auf diese abgestimmt werden.

Information Retrieval bezeichnet das gezielte Suchen nach und das (Wieder)finden von erschlossenen und gespeicherten Informationen anhand bestimmter Strategien unter Zuhilfenahme technischer Hilfsmittel. Diese Hilfsmittel werden als Retrievalsysteme bezeichnet und stehen zwischen dem Nutzer und den gespeicherten Informationen. Über Schnittstellen zu den Dokumenten und zu den Benutzern ermöglichen sie die Formulierung und Modifikation von Anfragen sowie die Darstellung von Treffermengen und bieten somit einen Zugang zu den in der Kollektion enthaltenen Dokumenten und den darin enthaltenen Informationen. In diesem Vermittlungsprozess zwischen Dokumentinhalten und den Nutzern mit ihren Informationsbedarfen kommt der Dokumentationssprache folgende Aufgabe zu:

Als wichtiges Werkzeug und begriffliche Grundlage des Information Retrievals stehen Dokumentationssprachen an der Schnittstelle zwischen der inhaltlichen Erschließung der Dokumente durch den Indexierer und den Anfragen des Benutzers, der in der Menge der mit Hilfe eben dieser Dokumentationssprache erschlossenen Dokumente die für seine Fragestellung relevanten Elemente sucht. Eine Aufgabe der Dokumentationssprache ist es dabei, die Vielgestaltigkeit und Komplexität der natürlichen Sprache wie sie in den Dokumenten vorkommt, auf ein einheitliches, kontrolliertes Vokabular zurückzuführen. Nur durch eine solche Reduktion und Normierung des verwendeten Vokabulars ist die, für eine funktionierende Information Retrieval notwendige Übereinstimmung von Erschließungs- und Suchvokabular zu gewährleisten.

Die Dokumentationssprache stellt somit die Grundlage für die intellektuelle, dokumentarische Erschließung und Bearbeitung der Dokumente sowie für die Formulierung der Suchanfragen durch den Benutzer und den automatischen Abgleich dieser beiden Aussagen dar. Beide Aspekte stellen besondere Anforderungen an die Gestaltung der Dokumentationssprache und an die Struktur des verwendeten Vokabulars.

Erschließung und Indexierung benötigen ein System, um Inhalte möglichst detailliert und umfassend darzustellen und in idealerweise koextensiven Indexaten zu fassen. Auf Seiten der Suchenden ist es entscheidend, dass abgeschätzt werden kann, mit welchen Elementen bestimmte Inhalte dargestellt werden. Das verwendete System muss also Wiedergabetreue, Ausdruckstärke und Vorhersagbarkeit vereinen.

Dies wird durch verschiedene Ansätze gewährleistet. Zum einen durch eine strikte Kontrolle des Vokabulars, zum anderen durch die präzise Formulierung der erlaubten Syntax. Für die Benutzung des Systems ist die Zugänglichkeit des Vokabulars für den Benutzer zu gewährleisten. Ebenso muss die Gesamtheit aller syntaktischen Elemente klar definiert und ihre Funktionen präzise und verständlich dargestellt werden.

Im Folgenden sollen die allgemeinen Eigenschaften des Vokabulars genauer dargestellt werden.

Grundlage jeder Dokumentationssprache ist ein kontrolliertes Vokabular sowie eine Syntax die die Verwendung und die möglichen Kombinationen der einzelnen Elemente dieses Vokabulars eindeutig regelt. Dabei kann es sich, wie im Falle von Klassifikationen um Klassen oder Elementarklassen handeln, oder um Begriffe, die durch natürlichsprachliche Bezeichnungen dargestellt werden.

Eine wichtige Funktion einer jeden Dokumentationssprache ist die Begrenzung und Kontrolle der Menge aller zu verwendenden Elemente. Hierbei kann es sich um Klassen oder Deskriptoren handeln, beides Repräsentanten präzise definierter Begriffe. Diese Reduktion des natürlichsprachlichen Ausgangsmaterials, und die Begrenzung und Definition des gesamten Vokabulars bezeichnet man als terminologische Kontrolle.

Die terminologische Kontrolle dient dazu, die potenziell unendliche Vielgestaltigkeit der natürlichen Sprache auf ein begrenztes Inventar an Begriffen zurückzuführen. Dabei arbeitet dieses System auf der Begriffsebene: Die enthaltenen Elemente sowie die zwischen ihnen bestehenden semantischen Relationen stellen das zwischen diesen Begriffen bestehende Beziehungssystem dar. Die vielfältigen möglichen Bezeichnungen für einen Begriff sind auch Teil dieses Systems, sie stellen den Übergang von der potenziell vielgestaltigen Bezeichnungsebene zur Begriffsebene dar und dienen der Bereitstellung weiterer Zugangsmöglichkeiten zum kontrollierten Vokabular; sie sind mit Elementen des kontrollierten Vokabulars durch Äquivalenzrelationen verbunden.

Neben der terminologischen Normierung auf der Bezeichnungsebene ist auch die begriffliche Strukturiertheit ein wesentliches Merkmal von Dokumentationssprachen. Eine möglichst ausgeprägte Strukturierung durch eine differenzierte Relationierung hat mehrere Funktionen. Sie ermöglicht eine übersichtliche Gestaltung des Vokabulars und dient der Orientierung des Benutzers. Entlang der durch Begriffsbeziehungen bestehenden Strukturen ist eine zielgerichtete Navigation und somit ein idealerweise intuitives Bewegen im begrifflichen Raum möglich.

Solche Strukturen können auch die Grundlage für komplexere Suchanfragen oder Suchstrategien sein. Um solche Funktionalitäten in einem Retrievalsystem zu implementieren, müssen die Strukturen des Vokabulars sowie der gebrauchten Syntax in entsprechende graphische Benutzeroberflächen mit den nötigen Steuerungs- und Suchfunktionen umgesetzt werden. Verschiedene Methoden der Visualisierung ermöglichen dabei die Abbildung eines oft sehr umfangreichen, komplexen und oftmals auch multidimensionalen abstrakten Begriffssystems auf einer graphischen Oberfläche in Form unterschiedlicher Strukturen. auf die zahlreichen Möglichkeiten und Verfahren der Visualisierung soll in diesem Zusammenhang nur am Rande eingegangen werden.

Eine weitere wichtige Funktion der Strukturierung des Begriffsraums anhand semantischer Relationen ist die Explikation zusätzlicher Informationen innerhalb eines Vokabulars einer Dokumentationssprache.

Anhand der Darstellung von Bezügen und Zusammenhängen, können Informationen, die in den einzelnen Elementen als solche nicht explizit enthalten sind und daher nicht dargestellt sind, für den Benutzer aber durchaus von Interesse sein können, in das Vokabular integriert werden. Diese zusätzlichen Informationen können vom Retrievalsystem zu Realisation zahlreicher Funktionalitäten genutzt werden. So können z.B. Suchanfragen entlang bestehender Hierarchieleitern erweitert, oder Suchterme durch Elemente aus ihrem semantischen Umfeld ergänzt oder unter Verwendung assoziativer Relationen modifiziert werden.¹² Bei ausreichender semantischer Anreicherung und dem Vorhandensein des entsprechenden notwendigen Formalisierungsgrads besteht die Möglichkeit, anhand entsprechender Systeme hochkomplexe Anfragen, unabhängig von Dokumentinhalten alleine aus den Strukturen des Vokabulars zu beantworten. Inwieweit diese Verfahren der Wissensrepräsentation durch semantische Technologien für das Information Retrieval genutzt werden können, soll in Kapitel 4.2 ausführlich dargestellt werden.

Während bei Klassifikationen die Einteilung der Begriffe in hierarchisch zueinander relationierte Klassen im Vordergrund steht, werden bei Thesauri die Begriffe direkt miteinander verknüpft und die zwischen ihnen bestehenden Beziehungen durch ein System an Begriffsrelationen ausgewiesen.

Neben der Strukturierung von Begriffsräumen bietet die Verwendung präzise definierter semantischer Relationen auch die Möglichkeit einer zusätzlichen relationalen Definition von Begriffen. Solche relationalen Definitionen, bei denen die Bedeutung eines Begriffs durch seine Beziehungen zu anderen Begriffen festgeschrieben werden kann, können neben extensionalen oder intensionalen Definitionen zusätzliche Bedeutungselemente in einen Begriff einschreiben und diesen somit präzisieren. So können durch Relationen ganze Begriffssysteme modelliert und damit Aussagen über die „Gestalt“ des dargestellten Gegenstandsbereichs getroffen werden, die über die in den Definitionen der einzelnen Begriffe enthaltenen Informationen hinausgehen.

Dies ist vor allem dann von Interesse, wenn wie im Fall der SWD die einzelnen Begriffe nicht durch eine eigene Definition erklärt werden, sondern aufgrund ihrer Gebräuchlichkeit als selbsterklärend empfunden oder als bekannt vorausgesetzt werden. Auch wenn relationale Definitionen oder durch Relationen dargestellte Informationen eine eigenständige natürlichsprachliche intensionale Definition nicht ersetzen sondern nur ergänzen können, haben sie für die Dokumentationssprache in ihrer Gesamtheit eine große Bedeutung. Fehlen diese intensionalen Definitionen oder andere Erklärungen des einzelnen Begriffs wird noch deutlicher „dass sich die Kon-

¹² s. a. Kapitel 4 dieser Arbeit

zeptualisierung eines Sachverhalts nie allein auf einen Begriff begrenzt, sondern immer auch Teile des semantischen Umfelds einbezieht“¹³.

Allerdings werden die Möglichkeiten der Bedeutungsdefinition und der Erklärung von Bedeutung durch die Abbildung relationaler Zusammenhänge sehr unterschiedlich bewertet, zumal wenn es sich um traditionelle Typen von Dokumentationssprachen mit einem sehr begrenzten Inventar an Relationstypen handelt. Gegen den Ansatz einer relationalen Begriffsdefinition wird auch die Schwierigkeit angeführt, natürlichsprachliche und damit in ihrer Entstehung und ihrem Gebrauch nicht nach logischen Prinzipien aufgebaute Begriffe ausschließlich durch ihre Verknüpfung mit anderen, ebenfalls nicht definierten und bestenfalls wiederum anhand von Relationen eingegrenzten Begriffen zu erklären oder zumindest semantisch grob einzugrenzen.

Während Larivière die Möglichkeiten einer präzisen relationalen Definition generell in Frage stellt,¹⁴ betonen andere die Unzulänglichkeit eines auf drei allgemeine und nicht präzise definierte Relationen reduzierten Relationspektrums.¹⁵

Diese Einwände sind zwar berechtigt, es bleibt aber zu beachten, dass die verwendeten Relationen nicht primär als Instrumente zur Definition der einzelnen Begriffe angesehen sondern als bewusst zur Schaffung anwendungsorientiert gestalteter und eingesetzter Strukturen verstanden werden: Deren Zweck ist die Unterstützung von Rechercheprozessen im Information Retrieval und nicht die Abbildung einer sprachunabhängigen, von Aspekten der Kommunikation und Suche abstrahierten Wirklichkeit. Die präzise Definition der einzelnen Begriffe sowie die Abbildung von deren semantischem Gehalt und logischen Eigenschaften ist für die Realisation solcher Funktionalitäten nicht zwingend notwendig und überdies nicht zu leisten; die Interpretation des einzelnen Begriffs ist intellektuell zu leisten. Insofern ist die angestrebte semantische Anreicherung als eine Präzisierung auf struktureller nicht auf begrifflicher Ebene zu sehen.

2.4 Strukturen und Typen von Dokumentationssprachen

Neben der Kontrolle des Vokabulars haben Dokumentationssprachen auch die Aufgabe, das zu verwendende Begriffsinventar zu strukturieren. Dass Dokumentationssprachen und Vokabularien unabhängig von Umfang, Komplexität und semantischer Aussagekraft immer als strukturierte Systeme in ihrer Gesamtheit gesehen werden müssen, ergibt sich aus der Bedeu-

¹³ Gödert, Winfried; Klaus Lepsky. Semantische Umfeldsuche im Information Retrieval in Online-Katalogen.

Köln: Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaften, Institut für Informationswissenschaft 1998. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft 7) S. 13.

¹⁴ Vgl. Larivière, L. Vers un produit unifié en terminologie et en documentation: le thésaurus terminologique. *Méta*, 34 (1989) S. 457-467

¹⁵ Vgl. Sager, J. C. A practical course in terminology processing. Amsterdam: Benjamins. 1990.

tung, die der oben beschriebenen Vermittlung kognitiver Strukturen zukommt.

Die Strukturierung der Dokumentationssprache kann durch verschiedene Verfahren, welche auch mit einander kombiniert werden können, erreicht werden. Die Ordnung aller Begriffe in Klassen eines Klassifikationssystems oder die Strukturierung eines Vokabulars durch die Ausweisung unterschiedlicher Relationen zwischen den Begriffen finden in verschiedenen Typen von Dokumentationssprachen ihre Anwendung. Im Folgenden sollen die wichtigsten dieser Typen kurz vorgestellt werden und vor allem in Hinblick auf die in ihnen verwendeten Relationstypen analysiert werden. Hierbei soll untersucht werden, welche Arten von Relationen in den verschiedenen Arten verwendet werden und wie sich dies auf die Funktionsweise und Handhabbarkeit dieser Dokumentationssprachen auswirkt.

Dabei sollen die Typen Thesauri und semantische Netze in dieser übersichtartigen Darstellung nur kurz behandelt werden, da ihnen jeweils eigene Kapitel (Kapitel 3 für Thesauri, Kapitel 4 für semantische Netze) für eine ausführlichere Darstellung zugeordnet sind.

Diese oben dargestellte Strukturierung des Vokabulars hat mehrere Funktionen:

Zum einen ermöglicht sie eine übersichtlichere Präsentation des gesamten Begriffssystems und ist Grundlage für verschiedene Möglichkeiten der Visualisierung. Zum anderen dient die Konstruktion und Benennung von Klassen und der Ausweis von Begriffsrelationen der Unterstützung der Definition von Begriffen. Der weitaus wichtigste Aspekt ist allerdings die Schaffung von Strukturen, welche nicht die Definition der einzelnen Begriffe anstreben, sondern Bezüge und Zusammenhänge innerhalb des Vokabulars in seiner Gesamtheit aufzeigen und nutzbar machen wollen.

Nach einem Überblick über verschiedene Arten von Dokumentations-sprachen sollen grundlegende Prinzipien und Gestaltungsmöglichkeiten dieser Dokumentationssprachen unter besonderer Berücksichtigung ihrer Relationierung und ihrer Funktionsprinzipien im Retrieval erörtert werden. Neben den unterschiedlichen Funktionsweisen sollen auch die verschiedenen Grade der semantischen Relationierung sowie die hierzu verwandten Relationstypen innerhalb dieser verschiedenen Typen untersucht werden. Die zunehmend umfangreiche und differenzierte Relationierung der dargestellten Arten von Dokumentationssprachen kann als eine kontinuierliche Entwicklung hin zu einer immer größeren „semantischen Reichhaltigkeit“¹⁶ verstanden werden, an deren Ende Ontologien als die begriffliche Grundlage von semantischen Technologien und Strukturen zur Wissensorganisation und möglicherweise als Hilfsmittel zur Erschließung stehen.

¹⁶ Pellegrini, Tassilo; Andreas Blumauer Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen . In: Pellegrini, Tassilo; Andreas Blumauer (Hrsg.) Semantic Web. Wege zur vernetzten Wissensgesellschaft. Berlin [u.a.] Springer, 2006 S. 9-24. hier S. 16

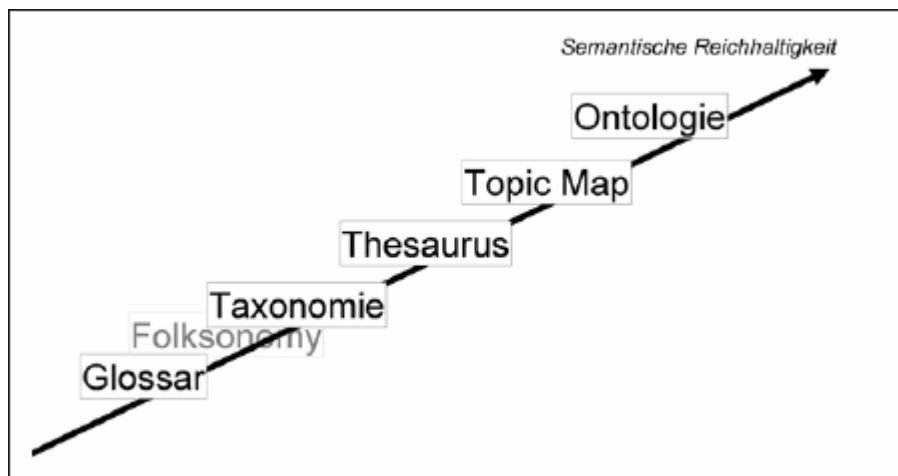


Abbildung 2: „semantische Treppe“¹⁷

Hierbei ist vor allem die Weiterentwicklung von Thesauri hin zu Ontologien und semantischen Netzen von besonderem Interesse. Um die Zusammenhänge und Unterschiede zwischen diesen Typen von Dokumentationssprachen bezüglich ihrer Relationierung herauszuarbeiten, sollen im Folgenden die gängigsten Typen dargestellt werden. Dies soll allerdings keine allgemeine Darstellung sein, sondern sich auf die für die Fragestellung dieser Arbeit relevanten Aspekte der Relationierung sowie der jeweils spezifischen Relationsarten konzentrieren.

2.4.1 Klassifikationen

Anders als Thesauri und Schlagwortlisten dienen Klassifikationen nicht der verbalen Erschließung sondern ordnen Dokumente oder Informationen durch Bereitstellung zahlreicher, nach bestimmten Kriterien geordneter Klassen, welche in einer hierarchischen Gesamtstruktur zueinander geordnet sind. Dieser klassifikatorische Ansatz scheint auf den ersten Blick mit der Idee einer zur Erschließung verwendeten „Sprache“ und einem dazu notwendigen Vokabular nicht viel gemein zu haben, dennoch sollen hier Klassifikationen als Sonderform der Indexierungssprachen kurz erörtert werden. Vor allem die Verfahren der Facettenklassifikation mit ihren verschiedenen Möglichkeiten, Instanzen von Elementarklassen zu komplexeren Begriffen zusammenzufügen, können für die nachfolgenden Überlegungen zur Gestaltung von Dokumentationssprachen sowie deren Relationierung von Interesse sein. Des Weiteren sind Klassifikationen auch für Hybridformen aus ver-

¹⁷ Die Grafik ist der folgenden Quelle entnommen: Pellegrini, Tassilo; Andreas Blumauer Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen . In: Pellegrini, Tassilo; Andreas Blumauer (Hrsg.) Semantic Web. Wege zur vernetzten Wissensgesellschaft. Berlin [u.a.] Springer, 2006 S. 9-24. hier S. 16

schiedenen Arten von Dokumentationssprachen und Erschließungsansätzen von Bedeutung.¹⁸

Klassifikationen ermöglichen die Ordnung von Informationen durch die Definition von Klassen und deren Ordnung in einem hierarchischen Gerüst sowie durch die Bezeichnung der einzelnen Klassen durch Notationen. Eine wesentliche typologische Unterscheidung ist die zwischen präkombinierten und facettierten Klassifikationssystemen. Beide ermöglichen die Darstellung komplexer Begriffe sowie deren Verhältnis zueinander in unterschiedlicher Ausprägung. Während bei der Methode der Präkombination die Kombination mehrerer Aspekte innerhalb einer monolithischen Klasse natürlichsprachlich dargestellt wird, und nachträglich mit einer Notation repräsentiert wird, ist die Kombination mehrerer Elementarklassen verschiedener Facetten zu einem komplexen Begriff nach genau definierten Regeln bei einem facettierten Verfahren auch in der Notation zu erkennen.

2.4.1.1 Präkombinierte Klassifikationssysteme

Bei präkombinierten Klassifikationen hat die Struktur, die durch die Definition von Beziehungen zwischen einzelnen Klassen in Form von Ober- und Unterklassen erschaffen wird und durch die Notation wiedergegeben werden kann, meistens einen deutlich hierarchischen Charakter. Die Verwendung einer strukturabbildenden Notation bei einer präkombinierten Klassifikation verdeutlicht die hierarchische Struktur. Sie dient der Orientierung des Benutzers und als Grundlage verschiedener Funktionalitäten des Information Retrieval wie zum Beispiel die Möglichkeiten einer Notationstrunkierung. Die hierarchischen Relationen, welche die einzelnen Klassen verbinden und so den strukturellen Zusammenhang erzeugen, bestehen zwischen den Klassen. Beziehungen zwischen Begriffen die Teile von Klassen sind, bestehen zwar auf der natürlichsprachlichen Ebene, können aber aufgrund der Abgeschlossenheit der präkombinierten Klassen nicht differenziert ausgewiesen werden. Ebenso werden assoziative Relationen zwischen verwandten Klassen nur auf der Klassenebene dargestellt. Auf eine semantische Differenzierung unterschiedlicher Typen assoziativer Relationen wird in den meisten Fällen verzichtet.

Äquivalenzrelationen, die zwischen mehreren Synonymen bestehen und auf eine Vorzugsbenennung hinweisen, werden in den „Tafeln“ nicht dargestellt sondern sind in ein Register ausgelagert. Somit sind auch Synonyme Teil der Gesamtstruktur, auch wenn sie in ihrer hinführenden Funktion außerhalb des Begriffinventars zu stehen scheinen.

2.4.1.2 Facettenklassifikationen

Anders als präkombinierte Klassifikationen verwenden Facettenklassifikationen ein genau definiertes Inventar an Elementarklassen zur Erstellung

¹⁸ s. a. Kapitel 2.4.3. dieser Arbeit

komplexer Begriffe und der entsprechenden Notationen durch Addition und Synthese.

Für die Frage einer möglichen Weiterentwicklung sind folgende Eigenschaften von Facettenklassifikationen oder anderer facettierter Systeme interessant: Das Bestehen eines Inventars von Primitiven oder Elementarbegriffen, die das zu verwendende Ausgangsvokabular darstellen sowie das Vorhandensein genauer syntaktischer Regeln zur Synthetisierung der komplexen Begriffe. Die darin enthaltenen syntaktischen Relationen können unter Umständen auch Orientierung geben für Relationen, die in anderen Begriffsinventaren als semantische Relationen vorhanden sein können.¹⁹

Ähnlich verhält es sich mit Rollenoperatoren welche eine ausdrucksstarke syntaktische Indexierung ermöglichen und die Darstellung komplexer multidimensionaler Sachverhalte durch die Ausweisung zahlreicher unterschiedlicher Relationen erlaubt. Der semantische Gehalt dieser Rollenoperatoren sowie die durch diese Operatoren ausgedrückten Beziehungen zwischen einzelnen Begriffen oder Elementen können zur Konzeption semantischer Relationen herangezogen werden. Solche an sich syntaktischen Relationen können auch als semantische Relationen in Begriffssystemen funktionieren.

2.4.2 Thesauri

Thesauri stellen eine in verschiedenen Bereichen bereits weit entwickelte Form von Dokumentationssprachen mit vielfältigen Ausprägungen und Einsatzmöglichkeiten dar.

Dies soll in Kapitel 3 ausführlicher behandelt werden, um ausgehend von klassischen Thesaurusformen aufzuzeigen, welche Entwicklungs- und Gestaltungsmöglichkeiten in einer differenziert ausgearbeiteten und strukturiert gestalteten Relationierung liegen.

2.4.3 Hybridformen

In Fällen in denen die zuvor beschriebenen Verfahren der Strukturierung nicht ausreichen oder aus verschiedenen Gründen nicht hinreichend umgesetzt werden konnten, gibt es Überlegungen, unterschiedliche Verfahren zu kombinieren oder bestehende Strukturen durch die Integration zusätzlicher Merkmale zu ergänzen. So können zum Beispiel schwach relationierte Vokabularen und Thesauri durch die systematische Verknüpfung aller Begriffe mit Elementen anderer, auch andersartig oder umfangreicher strukturierter Systeme wie sie zum Beispiel Klassifikationen darstellen, ähnlich einer Konkordanz, an Strukturiertheit gewinnen und somit für zusätzliche Funktionen genutzt werden. Im Unterschied zu den, in zahlreichen Thesauri vergebenen Notationen oder Systemstellen wird bei Hybridformen versucht,

¹⁹ Zum Zusammenhang syntagmatischer und paradigmatischer Relationen in Dokumentationssprachen s.a. Green, Rebecca.: Syntagmatic relationships in index languages: a reassessment. In: Library quarterly. 65 (1995) (4). S. 365-385.

alle Begriffe in einem Vokabular mit möglichst entsprechenden Klassen und den in ihnen enthaltenen Begriffen zu verknüpfen, wobei der Grad der Übereinstimmung zwischen den Thesaurusbegriffen und den zugeordneten Klassen durch „Bindungsstärken“ dokumentiert werden kann. So werden im Projekt CrissCross²⁰ die Schlagwörter der Schalgwort-Normdatei mit den entsprechenden Notationen der Dewey Decimal Classification (DDC) verknüpft, um damit die Vergabe und den Gebrauch von DDC Notationen während der Erschließung zu erleichtern und die Nutzung dieser strukturabbildenden Notationen für die Recherche zu ermöglichen. Der damit einhergehende Gewinn an dringend benötigter Strukturiertheit ist offensichtlich und macht darüber hinaus die Notwendigkeit einer ausführlicheren, umfangreicheren und umfassenderen Relationierung deutlich.

Viele dieser Hybridformen haben einen experimentellen Charakter und werden oft aus bestehenden Anwendungen heraus entwickelt. Dabei geht es um die Verbesserung bestimmter Funktionalitäten einzelner Anwendungen bereits bestehender und oftmals bewährter Dokumentationssprachen.

Auch die Überlegungen dieser Arbeit zur semantischen Anreicherung sowie der Integration semantischer Technologien haben Versuchscharakter. Hierbei geht es nicht darum, semantische Netze, wie sie in bereits weit entwickelten und hochspezialisierten Systemen zur Diagnostik oder Simulation im Bereich der Medizin oder der Verfahrenstechnik existieren, für den Bereich der Dokumentationssprachen nachzuahmen oder zu simulieren. Vielmehr sollen einzelne Aspekte semantischer Technologien herausgegriffen werden und für die Gestaltung von Dokumentationssprachen und das Information Retrieval nutzbar gemacht werden. Dabei werden vor allem die Möglichkeiten der Definition, Typisierung und Konstruktion von Relationen und ihre Verwendung zur Modellierung komplexer begrifflicher Strukturen von Interesse sein. Daher sollen im folgenden Kapitel die Besonderheiten, Funktionsweisen und Gestaltungsmöglichkeiten von semantischen Netzen erörtert werden. Hierbei sollen vor allem die Aspekte der Relationen sowie deren Funktionen im Vordergrund stehen, wie sie dann im fünften Kapitel auf die Gestaltung eines konkreten Gegenstandsbereichs angewandt werden soll.

2.4.4 Semantische Netze

Semantische Netze sind Repräsentationsformalismen welche das Wissen über einen bestimmten begrenzten Gegenstand durch die präzise Definition von Begriffen sowie die Ausweisung der zwischen diesen Begriffen bestehenden Beziehungen modellieren und somit für die maschinelle und kognitive Verarbeitung zur Verfügung stellen. Eigenschaften und Funktionsweisen semantischer Netze sollen in Kapitel 4 ausführlich behandelt werden.

²⁰ zum Projekt „CrissCross“ s. a. die Homepage des Projekts unter <http://www.fbi.fh-koeln.de/institut/projekte/CrissCross/index.html>

3 THESAURI

Grundprinzip des Thesaurus im informationswissenschaftlichen Sinne ist die Kontrolle und Strukturierung eines begrenzten Vokabulars zu Erschließungszwecken.

Diese für einen Thesaurus wichtige Funktion der Kontrolle sowie der Strukturierung des Vokabulars wird durch die Verwendung referentieller und relationaler semantischer Relationen erreicht.²¹ Referentielle semantische Relationen, ermöglichen es, die Vielgestaltigkeit und Reichhaltigkeit der natürlichen Sprache in ein kontrolliertes Vokabular zu überführen, indem sie von verschiedenen Bezeichnungen für bestimmte Begriffe auf eine einzige, normierte Vorzugsbenennung verweisen.

Zusammenhänge und Bezüge zwischen den Elementen dieses kontrollierten Vokabulars können durch eine relationale Semantik dargestellt werden.²² Eine solche relational definierte Semantik kann durch die zusätzliche Ausweisung weiterer semantischer Relationen differenziert gestaltet werden. Um eine solche Semantik aussagekräftig und idealerweise transparent gestalten zu können bedarf es der präzisen Definition aller verwendeter Relationstypen. In klassischen Thesauri wird diese Semantik nur anhand einiger weniger Relationstypen sowohl referentieller als auch relationaler Art dargestellt. Im folgenden Unterkapitel sollen die Eigenschaften dieser herkömmlichen Relationstypen erörtert werden, bevor im darauf folgenden Abschnitt über die verschiedenen Überlegungen und Ansätze zu einer Differenzierung und Verfeinerung der semantischen Relationen gesprochen werden soll.

Während der Aspekt der terminologischen Kontrolle sowie der damit verbundenen begrifflichen Normung allen Thesauri gemein ist, können sich Thesauri vor allem im Hinblick auf die Art und den Differenzierungsgrad ihrer Strukturierung erheblich voneinander unterscheiden. Dieser Differenzierungsgrad hängt von verschiedenen Faktoren ab und bestimmt im Wesentlichen die Möglichkeiten der Einbindung solcher Thesauri in Information Retrieval Systeme und deren Verwendung für die Recherche.

Kernstück aller herkömmlicher Thesauri ist das mehrstufige Etageengerüst, welches durch die Verwendung einfacher hierarchischer Relationen und der entsprechenden Begriffsbeziehungen von Ober- und Unterbegriff zwischen Begriffen gebildet wird. Die durch die vielfache Kombination zahlreicher Verbindungen von je zwei Begriffen durch die Ausweisung jeweils entsprechender Ober- und Unterbegriffe entstehenden Begriffsleitern

²¹ Zur Unterscheidung referentieller und relationaler Semantik vgl. Mazzocchi, Fulvio; Melissa Tiberi, Barbara De Santis, Paolo Plini. Relational semantics in thesauri: some remarks at theoretical and practical levels. In: Knowledge Organization Vol. 34 (2007) 4 S. 197-214, hier S. 199

²² Vgl. Swenonius, Elaine. The intellectual foundation of information organization. Cambridge, MA: The MIT Press. (2000).

bilden das Grundgerüst, welches durch weitere Verknüpfungen zwischen Begriffen ergänzt werden kann. Die mono- oder polyhierarchischen Strukturen lassen sich übersichtlich in einfachen Baumstrukturen darstellen und ermöglichen ein intuitives Erfassen und Verstehen der Begriffszusammenhänge sowie der Gesamtheit des verwendeten Begriffssystems.

Die meisten Thesauri verfügen über ein Minimalinventar an semantischen Relationen welches eine einfache hierarchische Strukturierung des Vokabulars sowie die Umsetzung rudimentärer Verweisstrukturen zwischen Begriffen, welche nicht in einer hierarchischen Beziehung zueinander stehen, ermöglichen. Dieses Minimalinventar soll im Folgenden dargestellt werden. Hierbei sollen auch die Probleme welche sich aus solchen Minimalinventaren sowie bestimmten Praktiken bei der Definition und Verwendung unspezifischer Relationstypen ergeben können eingegangen werden.

3.1 Herkömmliche Relationstypen

Historisch gesehen sind die gebräuchlichen Begriffsbeziehungen in Thesauri hierarchische Relationen, assoziative Relationen und Äquivalenzrelationen. Diese Bezeichnungen und die damit einhergehende grobe Einteilung in diese drei allgemeinen Relationstypen, haben ihren Ursprung im traditionellen Verweissystem klassischer Zettelkataloge. Anfangs waren die in Zettelkästen und -katalogen gebräuchlichen Verweisarten, „siehe“, „siehe auch (einseitig)“ und „siehe auch (zweiseitig)“ eins zu eins in Begriffsrelationen für Thesauri umgesetzt worden.

So entstanden die drei oben genannten Relationstypen sowie die davon abgeleiteten Begriffsbeziehungen „Ober/Unterbegriff, (OB/UB) „verwandter Begriff“ (VB) und die zur Darstellung von Äquivalenzrelationen gebrauchte Anweisungen welche je nach Thesaurus unterschiedliche Bezeichnungen haben können. Die dafür verwendeten Bezeichnungen stellen gerade im Fall der Begriffe der „assoziativen Relation“ und des „verwandten Begriffs“ möglicherweise nur Verlegenheitslösungen dar. Der Versuch einer theoretischen Begründung und systematischen Reflexion der unterschiedlichen Relationstypen in Thesauri folgte erst später mit Wersig.²³ Hierbei wurde deutlich, wie wenig definiert und unpräzise die assoziativen Relationen in ihrer ganzen potenziellen Reichhaltigkeit waren. Dabei betont er, dass sich Thesauri gerade darin von traditionellen bibliothekarischen Schlagwortsystemen unterscheiden und auszeichnen, „dass sie Begriffsbeziehungen explizit und systematisch darstellen.“²⁴

Während der Begriff der hierarchischen Relation sowie deren mögliche Ausprägungen und deren Funktionsweise gut dokumentiert und abgesichert

²³ Wersig, Gernot. Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. München: Saur, 1985 (DGD-Schriftenreihe 8) hier besonders S. 111

²⁴ Wersig, Gernot. Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. München: Saur, 1985 (DGD-Schriftenreihe 8) S. 123

ist, ist der Begriff der assoziativen Relation kaum definiert.²⁵ Wie schwierig eine Definition assoziativer Relationen und eine angemessene Behandlung ihrer Vielgestaltigkeit ist, zeigt sich an den oft „ex negativo“ gestalteten Definitionen die diese oft nur gegenüber hierarchischen Relationen abgrenzen. Beispiele für solche Definitionen „ex negativo“ finden sich bei Aitchison²⁶ und auch bei Soergel, wird die assoziative Relation im Ausschlußverfahren definiert.²⁷

Dieses Verständnis dieser so genannten „assoziativen Relationen“ ist ebenso wie die dafür verwendete und mittlerweile etablierte Bezeichnung selbst problematisch. Sie sind gerade für den Gebrauch in weiter entwickelten elektronischen Systemen aber auch für den fachlichen Diskurs über Wesen, Funktion und den Gebrauch von Relationen für Aufgaben der Dokumentation und der Wissensorganisation nicht angemessen und erschweren ein differenzierteres Verständnis über semantische und logische Eigenschaften und des sich daraus ergebenden funktionalen Potenzial der zahlreichen unter diesem Begriff subsumierten Relationstypen.

Daher soll die Bezeichnung der „assoziativen Relation“ sowie die auf diese Relation zurückgehende Begriffsbeziehung des „verwandten Begriffs“ bei der Analyse und der Konzeption von Relationstypen und Begriffsbeziehungen in den folgenden Kapiteln dieser Arbeit bewusst vermieden werden. Lediglich bei der Darstellung dieses Relationstyps in bereits vorhandenen Dokumentationssprachen und Systeme werden die entsprechend vorgefundenen Begrifflichkeiten und Bezeichnungen verwendet. Der vorgeschlagene Begriff der „nichthierarchischen Relationen“ bietet zwar ebenfalls nur eine Definition „ex-negativo“, verdeutlicht dieses Problem jedoch und ermöglicht und fordert daher geradezu eine genauere Ausdifferenzierung und spezifischere Definition der darunter subsumierten Relationen.

Die beobachtete mangelnde semantische Differenzierung und Präzisierung resultiert teilweise auch aus der Praxis, dass bei der Definition von Begriffsbeziehungen nicht genau zwischen deren semantischem Gehalt und ihren logischen Eigenschaften unterschieden wird. Des weiteren wird eine notwendige theoretische und konzeptionelle Reflektion bezüglich der möglichen verschiedenen Typen auch dadurch erschwert, dass in der praktischen Thesaurusarbeit lange Zeit oftmals Relationen nach praktischen Notwendigkeiten definiert und benutzt wurden ohne dies hinreichend zu reflektieren.

Bereits bei der Konzeption konventioneller Thesauri wurde erkannt, dass durch eine Differenzierung der verwendeten Relationen sowie deren strengen Formalisierung erhebliche Vorteile und Funktionszugewinne zu erwarten sind. Vor allem eine genauere Untersuchung, Differenzierung und Aus-

²⁵ Clarke, Stella G. Dextre. Thesaural Relationships. In: Bean, Carol A. Rebecca Green (Hrsg.): Relationships in the Organisation of Knowledge. Dordrecht: Kluwer Academic Publishers. 2001 (Information Science and Knowledge Management 2). S.37-52

²⁶ Aitchison, Jean; Alan Gilchrist, David Bawden. Thesaurus construction and use: a practical manual. London: Aslib 1997. S. 58

²⁷ Soergel, Dagobert. Indexing languages and thesauri: construction and maintenance. Los Angeles: Melville Publishing Company, 1974 (Information Sciences Series) S. 107

typisierung der nicht weiter definierten nicht-hierarchischen Relationen scheint lohnend. Das Potenzial, welches diese nicht-hierarchischen Relationen für die Realisation verschiedene Retrievalfunktionalitäten und Recherchekonzepte haben wird auch von Tudhope²⁸ und anderen²⁹ erkannt, jedoch nicht ohne die Notwendigkeit einer weitergehenden Spezifizierung und Austypisierung zu betonen.

Wie auch bei der schon als problematisch erkannten Begriffsbeziehung „verwandter Begriff (VB)“ besteht auch bei der Begriffsbeziehung welche aufgrund einer hierarchischen Relation existieren, das Problem, dass oftmals scheinbar ähnliche aber eben doch verschiedene Relationen die sich in ihrem semantischen Gehalt teilweise erheblich voneinander unterscheiden, unter diesem Begriff subsumiert werden, und eine genauere Darstellung der semantischen und logischen Unterschiede unterbleibt. So wird oft nicht zwischen den generischen und partitiven Relationen unterschieden, oder bestimmte Aspekte, die bei der Definition der jeweiligen Unterbegriffe benutzt wurden, werden nicht angemessen berücksichtigt und nicht differenziert dargestellt. Dies scheint auf den ersten Blick unproblematisch, da die Verwendung dieser allgemeineren Relation nicht falsch sondern lediglich unspezifisch ist. Die unterschiedlichen Aspekte nach denen die jeweils spezifischeren Unterbegriffe gebildet und geordnet werden, können vom Benutzer intellektuell erkannt und entsprechend interpretiert werden. Diese Arbeit wird vom Benutzer meistens unbewusst geleistet, Die aus solchen Aspektwechseln entstehenden Inkonsistenzen und logischen Brüche sind bei genauer Analyse zwar durchaus erkennbar, werden aber in Kauf genommen, um somit die einfache und übersichtliche Darstellung verschiedenartiger Begriffsbeziehungen mithilfe eines einzigen Relationstyps, der hierarchischen Relation zu ermöglichen. Der verständliche Wunsch nach Schaffung eines einfachen und überschaubaren Relationeninventars scheint wichtiger als die semantisch und logisch präzise Definition spezifischer Relationen. Die sich daraus ergebenden Schwierigkeiten hinsichtlich des Retrievals werden größtenteils nicht ausreichend reflektiert.

Ein solcher Verzicht auf eine weitergehende Differenzierung der Relationstypen ermöglicht die Relationierung und somit Integration verschiedenster Begriffstypen in durchgängige Begriffsleitern und die Reduzierung der Zahl unterschiedlicher Stränge oder Dimensionen, erfordert allerdings die genaue Analyse und Interpretation durch den Benutzer. Eine maschinelle Verarbeitung oder gar die maschinelle Interpretation eines solchen Relationengefüges sind nicht möglich, da wichtige Informationen bezüglich des semantischen Gehalts sowie der logischen Eigenschaften fehlen oder zumindest nicht hinreichend expliziert sind, oder die oben dargestellten Inkonsistenzen zu zahlreichen falschen Schlüssen oder zumindest zweifelhaften

²⁸ Tudhope, Douglas; Harith Alani; Christopher Jones. "Augmenting thesaural relationships: Possibilities for Retrieval". In: *Journal of Digital Information*, 1 (8) (2001)

²⁹ Cunliffe, D.; C. Taylor; Douglas Tudhope, "Query-based navigation in semantically indexed hypermedia". In: *Proceedings of the 8th ACM Conference on Hypertext*. (1997) S. 87-95.

Aussagen führen würden. So kann zum Beispiel das Fehlen der Differenzierung zwischen generischer und partitiver Relation – beides hierarchische Relationen – und deren vereinfachte Darstellung durch eine Ober- / Unterbegriffs-Relation zu Fehlschlüssen führen und auch für Benutzer das Verständnis erschweren oder Missverständnisse provozieren. Ebenso kann die durch mangelhafte oder gänzlich fehlende Unterscheidung verursachte Vermischung von generischer Relation und Instanz-Relation zu falschen oder zumindest unklaren Aussagen führen.³⁰

Noch stärker ausgeprägt ist das Problem der mangelnden Differenzierung bei den nicht-hierarchischen Relationen, welche die oben beschriebene hierarchische Grundstruktur ergänzen können. Dieser kaum definierte Relationstyp ist aufgrund seiner Vagheit zwar universell einsetzbar und erlaubt die Auszeichnung zahlloser, unspezifischer Relationen, die in solchen Relationen enthaltene semantische Information ist allerdings äußerst gering und lässt sich wiederum nur durch intellektuell zu leistende Interpretation erschließen, die auf implizitem Weltwissen und Wissen um den Kontext basierten³¹ Solche Relationen sind bei moderatem und überlegtem Gebrauch in konventionellen Umgebungen sicherlich von Nutzen, Zur Modellierung komplexer, differenziert gestalteter und damit aussagekräftiger Begriffssysteme, welche als Ontologien die Grundlage für die maschinelle Unterstützung der Benutzer im Rechercheprozess in Information Retrieval Systemen gebraucht werden sollen, sind sie aufgrund ihrer Ungenauigkeit nicht geeignet.

Das grundlegende Problem, dass ein intuitives Verständnis auch komplexer und unzureichend dargestellter Begriffsbeziehungen, und Konzeptausagen, wie es Benutzer aufbringen können, nicht von einer maschinellen Verarbeitung geleistet werden kann, kann nur in Ansätzen gelöst werden. Ein solcher Ansatz erfordert eine genaue Analyse und präzise Definition aller unterschiedlichen Relationstypen mit all ihren Eigenschaften in einer für die maschinelle Verarbeitung und Interpretation geeigneten Form. Dies ist nur durch die Definition und Verwendung umfangreicher und präzise ausgezeichneter Relationeninventare möglich.

3.2 Nicht-konventionelle Relationstypen

Es wurden bereits zahlreiche Versuche unternommen, die Relationeninventare unterschiedlicher Thesauri zu vergrößern und durch „nicht-konventionelle“ Thesaurusrelationen zu ergänzen. Eine solche Erweiterung wurde in einigen Fällen als eine Ergänzung durch zusätzliche „Spezial-

³⁰ Vgl. das „Dachdeckerbeispiel“ bei Fischer, Dietrich H. From Thesauri towards ontologies? In: *Advances of Knowledge Organization*, vol. 6 (1998) S. 18-30. hier S. 22 ff

³¹ Wersig, Gernot. *Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis*. München: Saur, 1985 (DGD-Schriftenreihe 8).

Relationen“ verstanden, es wurde aber auch auf die Möglichkeit und Notwendigkeit einer Differenzierung bekannter Relationen hingewiesen.³²

Schon damals wurde eine solche Differenzierung als die Korrektur struktureller Defizite angesehen und mit den Erfordernissen einer aussagekräftigen und funktionalen Modellierung der Thesauri als Dokumentationssprachen begründet: „Ihre Berechtigung erhalten diese nicht-konventionellen Relationen durch Leistungsmängel ihrer „klassischen Verwandten“ in deren Gesellschaft sie sich auch stets befinden.“³³

Neben zahlreichen Beispielen bei denen spezifischere und in vielen Fällen aus praktischen Überlegungen oder einer fachspezifischen Anwendungsorientierung heraus entwickelte Relationstypen ihren „konventionellen“ Gegenstücken zur Seite gestellt wurden, stehen auch Überlegungen, umfangreiche Relationeninventare von Grund auf neu zu entwickeln. Häufig wird dabei versucht, Inventare von Begriffsbeziehungen zu entwickeln, welche eine hinreichende Allgemeinheit mit einer größtmöglichen Spezifität verbanden. Neben Versuchen möglichst viele verschiedene Relationstypen in umfassenden und umfangreichen Typologien zu identifizieren und aufzuzeichnen, verfolgen andere die Schaffung übersichtlicher anwendungsspezifischer Relationeninventare, welche zur Modellierung spezialisierter Vokabulare oder Terminologien gebraucht werden.³⁴

Die zentrale Funktion der Relationen für die Strukturierung des Thesaurus sowie für die Funktion eines solchen Thesaurus innerhalb eines Information Retrievalsystems verdeutlicht die Gemeinsamkeiten mit semantischen Netzen, auf die zu einem späteren Zeitpunkt noch eingegangen werden soll. Auch der Vergleich der Relationen eines konventionellen Thesaurus mit den durch Kanten eines Graphen dargestellten Relationen eines semantischen Netzes wie ihn Reimer anstellt,³⁵ verdeutlicht die enge Verwandtschaft der beiden Methoden sowie das Entwicklungspotential welches sich durch die Erweiterung bestehender Thesauri zu semantisch angereicherten Repräsentationsinstrumenten durch die Differenzierung des Relationeninventars bietet.

Die Tatsache, dass sich die beiden wichtigsten gestalterischen Elemente semantischer Netze, nämlich die Bildung von Klassen sowie die Zuordnung entsprechender Instanzen und präzise definierte Relationen zwischen diesen Instanzen auch in Klassifikationen und Thesauri wieder finden, macht deutlich, welches Potenzial semantische Netze für den Bereich der Wissensorganisation speziell für die Aufgabe der Gestaltung kontrollierter Vokabulare haben und welche Rolle auch die selektive Verwendung einzelner se-

³² Schöndorf, Peter. Nicht-konventionelle Thesaurusrelationen als Orientierungshilfe für Indexierung und Recherche – Analyse ausgewählter Beispiele. In: Nachrichten für Dokumentation 39 (1988) S. 231-244

³³ Schöndorf, Peter. Nicht-konventionelle Thesaurusrelationen als Orientierungshilfe für Indexierung und Recherche – Analyse ausgewählter Beispiele. In: Nachrichten für Dokumentation 39 (1988) S. 231-244

³⁴ Fischer, Karl-Heinz. Der BERLIN-THESAURUS als Beispiel für einen komplexen Regionalthesaurus. In: Nachrichten für Dokumentation 39 (1988) S. 245-252

³⁵ Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik) hier vor allem S. 152 -154

mantischer Technologien für die Gestaltung von Dokumentationssprachen spielen kann.

4 SEMANTISCHE NETZE IM INFORMATION RETRIEVAL

In diesem Kapitel soll nach einer ersten begrifflichen Klärung des Begriffs der semantischen Netze, versucht werden, einige der Eigenschaften und Funktionsweisen dieser Repräsentationsformen darzustellen. Des weiteren sollen verschiedene Arten von semantischen Netzen und deren Ontologien voneinander abgegrenzt und verglichen werden. Anschließend sollen Strukturmerkmale und Gestaltungsmöglichkeiten semantischer Technologien vorgestellt werden. Hierbei soll vor allem erörtert werden, welche Möglichkeiten und Methoden bei der Konzeption, Modellierung und Kodierung von Ontologien auf die Gestaltung von Dokumentationssprachen übertragen werden können, und welche Prinzipien für die vorliegende Aufgabe der Relationierung von Dokumentationssprachen sinnvoll genutzt werden können. Diese Überlegungen sind vor dem Hintergrund der skizzierten Aufgabe zu sehen und erheben keinen Anspruch auf Vollständigkeit.

So werden manche Eigenschaften, Charakteristika und Gestaltungsmöglichkeiten semantischer Netze lediglich cursorisch behandelt. Der Fokus liegt auf den Möglichkeiten der Klassenbildung und der Relationentypisierung. Vor allem die weitergehenden umfangreichen logischen und algorithmischen Grundlagen für die Modellierung und das Funktionieren semantischer Netze können nicht umfassend dargestellt werden.³⁶

Eine präzise Definition des Begriffs „semantischer Netze“ ist ebenso wie die des damit eng verbundenen Begriffs der „Ontologie“ schwierig, da eine Vielzahl von Ausprägungen und Varianten existiert, welche sich in ihrer Entstehung, der Komplexität ihrer Gestaltung, ihrer Funktion und auch in der Art ihrer technischen Umsetzung erheblich voneinander unterscheiden. Dies ist bei dem folgenden Versuch einer ersten begrifflichen Klärung zu beachten, welche in erster Linie grundlegende Eigenschaften und Funktionsprinzipien herausarbeiten möchte und verschiedene Elemente unterschiedlicher anderer Definitionen³⁷ zusammenfügt:

Semantische Netze sind nonlineare, Repräsentationsformate und Repräsentationsformalismen, welche vielgestaltige Beziehungen zwischen Konzepten auszeichnen und darstellen. Ihr Zweck ist die formalisierte Modellie-

³⁶ Für eine ausführliche Darstellung der logischen Grundlagen vgl. Schmid, Ute; Martin Christof Kindsmüller. Kognitive Modellierung. Eine Einführung in die logischen und algorithmischen Grundlagen. Heidelberg [u. a.]: Spektrum Akademischer Verlag, 1996 sowie Sowa, John F.; Principles of semantic networks. Explorations in the representation of knowledge. San Mateo (CA): Morgan Kaufmann Publishers; 1991

³⁷ Diese Definition enthält Elemente der Definitionen aus dem Glossar der Semantic Web Company (ehemals Semantic Web School) <http://www.semantic-web.at/1.32.catchword.36.semantisches-netz.htm> sowie von Antoniou, Grigoris, Frank van Harmelen. A semantic web primer. Cambridge, MA [u. a.] MIT Press 2004 und Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik)

rung und Darstellung von Wissens- oder Beriffsstrukturen welche die Ausführung logischer Operationen und somit die Interpretation dieser Strukturen erlaubt.

Heutzutage wird zudem die maschinelle Verarbeitbarkeit von derart modellierten Strukturen als wichtige Voraussetzung für deren Verwendung zur Unterstützung unterschiedlicher Anwendungen in elektronischen Umgebungen betont.

Die netzartige Struktur wird durch die Verwendung von Knoten und Kanten zur Darstellung von Konzepten und den zwischen ihnen bestehenden Beziehungen erreicht.

Vorläufer semantischer Netze finden sich in verschiedenen wissenschaftlichen Disziplinen, seinen Ursprung hat das Konzept im Bereich der Kognitionspsychologie und der Neurowissenschaften.³⁸ Für eine eingehendere Darstellung der historischen Entwicklungen und Zusammenhänge sei hier ebenfalls auf Reimer³⁹ verwiesen. An dieser Stelle sollen lediglich zwei Aspekte der unterschiedlichen historischen Entwicklungen herausgegriffen werden: Zum einen die Entwicklung semantischer Netze aus psychologischen Modellen des menschlichen Gedächtnisses sowie der darin existierenden nicht weiter typisierten assoziativen Relationen und der damit verbundenen Problematik der mangelnden Interpretierbarkeit, sowie der daran anschließenden Frage nach der Möglichkeit der differenzierten Gestaltung solcher semantischer Netze. Die Auszeichnung unterschiedlicher Eigenschaften semantischer Relationen und der damit einhergehende erwünschte Detaillierungsgrad der semantischen Netze werden oft über einfache natürlichsprachliche Beschreibungen angestrebt. Dabei wird vor allem eine nicht ausreichend spezifizierte oder gänzlich „fehlende“ Semantik zahlreicher nicht präzise typisierter Netze als zentrales Problem erkannt:

„Die häufig fehlende Semantik der Knoten und Kanten in einem semantischen Netzformalismus, bedeutet, dass damit erstellte Repräsentationen genau betrachtet nicht interpretierbar sind, auch wenn die Knoten- und Kantenbeschriftungen eine bestimmte Bedeutung suggerieren“⁴⁰

Die formal strenge Ausweisung des semantischen Gehalts aller verwendeter Elemente wird als Grundlage für den Aufbau und die detaillierte Modellierung semantischer Netze, welche nicht mehr ausschließlich auf eine nur intellektuell zu leistende Interpretation angewiesen sind, erkannt. Selbst

³⁸ vgl. Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik S. 75. für eine eingehendere Darstellung der historischen Entwicklungen und Zusammenhänge siehe auch

Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik S. 149 ff.

³⁹ Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik

⁴⁰ Reimer, Ulrich. Einführung in die Wissensrepräsentation. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik) S. 150

für eine rein kognitive Interpretation ist eine differenzierte und aussagekräftige Definition der Konzepte und Typisierung der Relationen notwendig, wenn deren Interpretation ausschließlich auf den in der Repräsentation enthaltenen Informationen beruhen soll und nicht durch implizites Wissen ergänzt oder unterstützt werden soll.

Die Möglichkeit zur maschinellen oder zumindest maschinengestützten Verarbeitung und Interpretation dieser Netze hängt von einer formal strengen und semantisch expressiven Modellierung dieser Konstrukte ab. Um die formale Strenge dieser Netze in standardisierter Form gewährleisten zu können, werden verschiedene Beschreibungsmodelle und Sprachen entwickelt, von denen einige inzwischen den Status von de facto Standards erlangt haben:

Neben sehr einfachen Beschreibungsmodellen wie dem „*Resource Description Framework*“ (RDF) oder dem „*Simple Knowledge Organization System*“ (SKOS) welche lediglich die Modellierung und Darstellung einfacher Zusammenhänge und Sachverhalte ermöglichen, gibt es auch weiter entwickelte Sprachen, welche die Grundprinzipien und Gestaltungsmöglichkeiten der einfacheren Beschreibungsmodelle aufgreifen und sukzessive erweitern um so die Gestaltung und Kodierung umfangreicher, differenziert ausgearbeiteter, und vor allem strukturierter Ontologien ermöglichen. Im Rahmen dieser Arbeit sind dies das Beschreibungsmodell „*Resource Description Framework*“ (RDF), das dazugehörige Schema „*Resource Description Framework Schema*“ (RDFS oder RDFSchemata) sowie die Modellierungssprache „*Web Ontology Language*“ (OWL).⁴¹

Anhand des grundlegenden Beschreibungsmodells (*description framework*) RDF können einfache Aussagen getroffen werden. Dazu werden zwei Elemente (*ressources*) mit einer Eigenschaft (*property*) verbunden. Die so entstandene Konstruktion ist mit einem einfachen Aussagesatz bestehend aus Subjekt, Prädikat und Objekt oder einem Subjektkomplement vergleichbar.

Dieser ersten Ebene übergeordnet ist ein Schema oder Repräsentationsmodell, in welchem alle Elemente, welche zur Bildung dieser einfachen Aussagen herangezogen werden können gleich einem Inventar definiert und eingeführt werden. Das hierzu verwendete und zum jeweiligen Beschreibungsmodell gehörende Schema RDFS bietet hierbei die Möglichkeit dieses Inventar durch Klassenbildung und der Instanziierung der verwendeten Ressourcen alle Ressourcen zu definieren und zu ordnen.

Auch alle als Eigenschaften bezeichneten Relationstypen werden in diesem Schema in ihrem semantischen Gehalt sowie ihren logischen Eigenschaften definiert. Zudem können über die Angaben der *Range* und der *Domain* die Kombinationsmöglichkeiten und somit die Verwendung im Modellierungsprozess definiert und eingeschränkt werden. Dabei bezeichnet

⁴¹ Zu den funktionalen Zusammenhängen zwischen den einzelnen Beschreibungsmodellen, Modellierungs- und Auszeichnungssprachen s. a. Antoniou, Grigoris, Frank van Harmelen. A semantic web primer. Cambridge, MA [u. a.] MIT Press 2004

die Angabe „*Domain*“ aus welchen Klassen die Ressourcen stammen dürfen die in einer Aussage welche mithilfe dieses Relationstyps gemacht wird als „Subjekt“ fungieren. Die Angabe der „*Range*“ definiert die erlaubten Klassen der „Objekte“.

Sowohl Klassen als auch Eigenschaften können in hierarchischen Relationen zueinander geordnet werden, was eine übersichtliche Strukturierung und eine logische Beziehung ermöglicht aufgrund derer die logischen und semantischen Charakteristika von Klassen oder Eigenschaften an deren jeweilige Unterklassen (*subclasses*) oder Unter-Eigenschaften (*subproperties*) vererbt werden. Gerade für den Aufbau umfangreicher Inventare mit zahlreichen Klassen und Eigenschaften ist die Möglichkeit der hierarchischen Strukturierung ein wichtiges Instrument der Ordnung und Modellierung. Für die Entwicklung eines hierarchisch strukturierten Inventars von Relationstypen welche durch die in einem solchen Schema definierten „Eigenschaften“ dargestellt werden, ist die Verwendung eines solchen Schemas, welches hierarchische Strukturen unterstützt, eine Notwendigkeit.

4.1 Gestaltungsprinzipien semantischer Netze

Jedes semantische Netz besteht aus Knoten und Kanten. Egal ob es sich um einfache graphische Darstellungen oder komplexe, in Auszeichnungssprachen kodierte Ontologien handelt, sind die zugrunde liegenden Konstruktionsprinzipien die gleichen: Die Schaffung von Zusammenhängen und Strukturen durch die Verbindung mehrerer Elemente durch Relationen. Grundlage hierfür ist die Schaffung zahlreicher ineinandergreifender Kombinationen nach dem immer gleichen Bauplan: die Verbindung zweier Elemente zu einem Tripel mittels einer Beziehung. Diese Dreiverbindungen bilden die Elementarbausteine für einfache Netze sowie umfangreiche und komplexe Ontologien. Diese Triple stellen einfache Aussagen über den darzustellenden Gegenstandsbereich dar und ermöglichen in ihrem Zusammenwirken die detaillierte Abbildung umfangreicher Welten.

Sowohl die Anzahl der verwendeten Elemente sowie auch die Anzahl der unterschiedlichen Relationstypen ist dabei theoretisch unbegrenzt. Jedoch müssen alle Elemente präzise definiert und in entsprechenden Repräsentationsmodellen eingeführt werden, in denen die genauen semantischen und logischen Charakteristika aller Elemente in standardisierter Form hinterlegt sind. Um diese Repräsentationsmodelle, welche das gesamte Inventar möglicher Elemente zur Konstruktion eines semantischen Netzes vorhalten, möglichst übersichtlich zu strukturieren gibt es zahlreiche unterschiedliche Gestaltungsmöglichkeiten.

Im Folgenden sollen lediglich die grundlegenden Gestaltungsmechanismen der Klassenbildung und der Relationentypisierung erst allgemein vorgestellt werden. Daran anschließend soll dargestellt werden, wie diese Mechanismen in den Beschreibungsmodelle und Modellierungssprachen welche auch zur Gestaltung der Beispielontologie benutzt werden sollen. Dabei soll

auch auf die Funktionszusammenhänge zwischen Beschreibungsmodellen, deren Weiterentwicklung und Präzisierung anhand dazugehöriger Schemata sowie der daraus entwickelten vollwertigen Modellierungssprachen genauer eingegangen werden.

4.1.1 Klassenbildung

Semantische Netze unterscheiden sich neben den unterschiedlichen Methoden und Formen ihrer technischen Umsetzung vor allem durch den Differenzierungsgrad der Gestaltungsmöglichkeiten die zu ihrer Modellierung angewandt werden können. Um diese Gestaltungsmöglichkeiten immer weiter zu entwickeln und somit den Anforderungen eines immer differenzierter funktionierenden Information Retrievals nachzukommen, wurden unterschiedliche Sprachen und Beschreibungsmodelle entwickelt und immer weiter verfeinert. Ein besonderes Merkmal weiter entwickelter Beschreibungsmodelle und Modellierungssprachen ist die Möglichkeit, die Gesamtheit der verwendeten Elemente durch die Definition von Klassen zu ordnen.

Hierbei soll betont werden, dass der Begriff der Klassenbildung sowie die Verwendung von Klassen sich von dem Verständnis von Klassen im Kontext von Klassifikationssystemen unterscheidet. Im Kontext semantischer Technologien wird die Möglichkeit der Klassenbildung zur Schaffung logisch weitestgehend genau definierter Elementarbegriffe welche dann ebenfalls durch weitere Prozesse der Klassenbildung zu komplexen Begriffen kombiniert werden können.

Anders als bei präkombinierten Klassen herkömmlicher Klassifikationssysteme entstehen komplexe Begriffe hier durch eine präzise und formal darstellbare logische Modellierung auf der Basis von Elementarbegriffen, so zum Beispiel durch mengenlogische Operationen oder die Verwendung von Kardinalitäten. Eine solche Methode der Begriffsbildung ist bei natürlichsprachlich basierten Vokabularen nicht möglich, weshalb dieses Gestaltungsinstrument in solchen Fällen nicht zur Modellierung von Begriffen genutzt werden kann.

4.1.2 Typisierung von Relationen

Werden die Beziehungen welche zwischen Konzepten bestehen, lediglich als assoziative Relationen im Sinne einer nicht weiter definierten oder typisierten Verbindung verstanden, ist der Informationsgehalt, welcher anhand einer solchen Relation modelliert werden kann minimal: Der semantische Gehalt ist lediglich der einer „irgendwie“ bestehenden Verbindung, die logischen Eigenschaften sind nicht bekannt. Auch wenn die Darstellung einer solchen Beziehung durch eine ungerichtete Kante möglicherweise eine symmetrische Beziehung suggeriert, ist dies nicht sicher aus der Repräsentation alleine ersichtlich. Die Ausdruckskraft und die Möglichkeit einer präzisen und korrekten Interpretation einer ausschließlich anhand solcher Relationstypen gebildeten Repräsentationen ist äußerst gering. Um den Detaillierungsgang von Repräsentationen und damit auch ihre Expressivität zu erhö-

hen, können Eigenschaften von Beziehungen angegeben werden und somit unterschiedliche Beziehungs- oder Relationstypen unterschieden werden. Eine solche Definition kann sowohl den semantischen Gehalt, also die Bedeutung als auch die logischen Eigenschaften der Relationsarten oder Typen angeben und festschreiben. Die semantischen und logischen Eigenschaften aller Relationstypen müssen für die Modellierung des semantischen Netzes sowie für dessen Interpretierbarkeit in einem Repräsentationsmodell genau hinterlegt sein. Diese verschiedenen Relationstypen gehören ebenso wie die zuvor beschriebenen Klassen und deren Instanzen zum „Vokabular“ das zur Beschreibung von Gegenstandsbereichen oder Ausschnitten daraus benutzt werden kann. Um ein solches Vokabular überschaubar und handhabbar zu halten empfiehlt es sich, das Inventar an Relationstypen zu begrenzen und möglichst allgemeingültig zu gestalten.

„In einem semantischen Netz sind möglichst nur solche Beziehungskanten zu verwenden, die durch ein Repräsentationskonstrukt im zugehörigen Repräsentationsmodell bereitgestellt werden oder die den Status von Grundvokabular besitzen. Ad hoc eingeführte, domänenspezifische Beziehungskanten, die zu speziell sind, als dass sie sich als Grundvokabular eignen [...] sollten nur dann verwendet werden, wenn der sonst höhere Detaillierungsgrad und die sonst bessere Vergleichbarkeit verschiedener Repräsentationsausschnitte [...] nicht benötigt werden.“⁴²

In dieser Forderung sind bereits zwei wesentliche Punkte, welche auch für die Verwendung semantischer Technologien für die Gestaltung von Dokumentationssprachen von Bedeutung sind enthalten: Die Frage nach einem hinreichend allgemeingültigen aber dennoch expressiven Relationeninventar („Grundvokabular“) sowie die Frage nach den Möglichkeiten der Integration hochgradig spezifischer Relationstypen für spezielle Anwendungsfälle. Diese Fragen sollen im folgenden fünften Kapitel näher erörtert werden.

Das gesamte Vokabular einer Dokumentationssprache muss mit allen enthaltenen Elementen in seinen semantischen und logischen Eigenschaften in einem zu dem jeweiligen semantischen Netz zugehörigen Repräsentationsmodell festgelegt werden.

Um solche umfangreichen und komplexen Modelle besser strukturieren und somit die Aussagekraft dieser Konstrukte zu erhöhen, können in weiter entwickelten Beschreibungsmodellen und darauf aufbauenden Sprachen Klassen zur Ordnung der verwendeten Objekte definiert und die zur Konstruktion von Triplen verwendeten Relationen präzise typisiert werden. Struktur und Funktionsweise lassen sich mit der Form einfacher Aussagesätze vergleichen: Diese enthalten einfache präpositionale Aussagen welche sich formal aus einem „Subjekt“ einem „Prädikat“ sowie einem „Objekt“

⁴² Reimer, Ulrich. *Einführung in die Wissensrepräsentation*. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik) S. 99

zusammensetzen. In den entsprechenden Repräsentationskonstrukten und Modellen werden die Funktionen von Subjekt und Objekt von Elementen oder Konzepten, dargestellt. Im Rahmen des hier verwendeten Beschreibungsmodell RDF/RDFS werden diese Elemente als „Ressourcen“ (*individuals*), die Relationen welche zur Verknüpfung verwendet werden als „Eigenschaften“ (*properties*) bezeichnet.

Welche Relationen notwendig oder sinnvoll für die Anreicherung und Restrukturierung von kontrollierten Vokabularen sind, hängt auch von den Funktionalitäten ab, welche für das Information Retrieval auf diesen Strukturen aufsetzen sollen. Diese Funktionen sollen im Folgenden übersichtsartig dargestellt werden.

4.2 Erweiterung von Retrievalfunktionalitäten durch semantische Technologien

Struktur und Funktionsweise einer Dokumentationssprache müssen sich immer an den Erfordernissen des Retrievals orientieren. Um zu verstehen, welchen Nutzen die Übertragung semantischer Technologien in den Bereich konventioneller Dokumentationssprachen sowie deren Ausbau zu semantisch angereicherten Strukturen und terminologischen Ontologien haben kann, soll im Folgenden verschiedene Retrievalstrategien welche auf semantischen Strukturen aufsetzen dargestellt werden.

Hierbei soll aufgezeigt werden, auf welche Weise die Initiierung, Veränderung und Erweiterung von Suchanfragen im Retrievalprozess vorgenommen werden können und welche Rolle strukturierte Dokumentationssprachen bei der Orientierung des Benutzers sowie bei seinem Bestreben einen möglicherweise anfangs noch diffusen Informationsbedarf zu präzisieren und zu verbalisieren und anschließend in Form einer präzise formulierten Suchanfrage an das Retrievalsystem zu übergeben. Für alle diese Aufgaben sind die Orientierung im Begriffsraum sowie verschiedene Möglichkeiten der semantischen Umfeldsuche eine wichtige Voraussetzung.⁴³ Im Folgenden sollen die wichtigsten Vorgehensweisen und Strategien zur Verbesserung von Rechercheanfragen und entsprechender Suchergebnisse im Retrievalprozess dargestellt werden.

4.2.1 Orientierungsfunktion („Query-clarification“)

Voraussetzung für den effektiven und effizienten Gebrauch von Suchwerkzeugen sowie der begrifflichen Strukturen auf welche diese Suchwerkzeuge aufsetzen ist das Wissen des Benutzers um diese Strukturen und Zusammenhänge. Aus der Symmetrie des Retrievalprozesses erklärt sich der

⁴³ Zur Notwendigkeit und Möglichkeiten der semantischen Umfeldsuche in schwach relationierten Vokabularen s. a. Gödert, Winfried; Klaus Lepsky. Semantische Umfeldsuche im Information Retrieval in Online-Katalogen. Köln: Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaften, Institut für Informationswissenschaft 1998. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft 7) hier besonders S. 12-15.

Wunsch, der Benutzer möge möglichst genau über die Eigenschaften und Strukturen des verwendeten Vokabulars bescheid wissen und die Suchanfragen sowie deren Modifikationen entsprechend dieser Strukturen gestalten. Eine solche Forderung erscheint unrealistisch, trotzdem sind eine zumindest flüchtige Kenntnis des Gegenstandsbereichs sowie eine Orientierung des Benutzers im verwendeten Vokabular unabdingbar für die sinnvolle Umsetzung der oben beschriebenen Strategien. Daher kommt dem Vokabular auch eine Orientierungsfunktion zu. Das Vokabular soll durch die Begriffe und Strukturen einen ersten groben Überblick über den Gegenstandsbereich bieten, dessen genauere explorative Erkundung ermöglichen und somit auch die Klärung und Präzisierung des Informationsbedarfs im Vorfeld der eigentlichen Recherche unterstützen.

Selbst wenn man doch davon ausgehen darf, dass der Benutzer sehr wohl eine mehr oder wenig präzise Vorstellung über die von ihm benötigten und gesuchten Informationen besitzt, ist der Prozess der Konzeptualisierung sowie der Verbalisierung dieses Informationsbedarfs mit Beginn der Recherche noch keineswegs abgeschlossen. Dieser Transformationsprozess von einem möglicherweise diffusen oder vagen Informationsbedarf hin zu einer präzise formulierten Suchanfrage ist ein komplexer interaktiver Prozess. Dabei kommt es darauf an, dem Benutzer sinnvolle Hilfestellungen anzubieten, die es ihm ermöglichen, das Vokabular und die Syntax der verwendeten Dokumentationssprache sinnvoll zu gebrauchen und so effizient für einen erfolgreichen Rechercheprozess einzusetzen.

Die Orientierung im Vokabular, die Sichtung und die Auswahl von Suchtermen sowie die Zusammenstellung geeigneter Suchanfragen liegen in der Verantwortung des Benutzers, allerdings ist er dabei durch geeignete konzeptionelle und technische Maßnahmen zu unterstützen.⁴⁴

4.2.2 Modifikation („Query-modification“)

Neben der initialen Formulierung einer ersten Anfrage kommt auch der Modifikation von Suchanfragen in Abhängigkeit zu den erzielten Ergebnissen eine besondere Bedeutung zu.

Eine solche Modifikation kann zum einen dazu gebraucht werden, bei zu großen Treffermengen die Suchanfrage zu präzisieren und somit die Präzision zu erhöhen, oder aber auch um durch eine Erweiterung der ursprünglichen Anfrage weitere Aspekte mit in die Suche einzubeziehen und die Gesamtmenge der Treffer zu erhöhen. Dies kann mit einer Erhöhung des Recalls einhergehen, ist aber nicht automatisch mit dieser gleichzusetzen.

Solche Modifikationen oder Erweiterungen können auf verschiedene Weisen vom System vorgenommen oder von diesem unterstützt werden. Wird eine solche Modifikation automatisch vom Retrievalsystem vorgenommen, werden die Suchparameter nach entsprechenden Vorlagen und

⁴⁴ Stock, Wolfgang G. Information Retrieval. Informationen suchen und finden. München: Oldenbourg, 2007. S. 476 ff

Algorithmen und unter Berücksichtigung der Suchresultate der ersten ursprünglichen Anfrage modifiziert. Hierbei ist eine Interaktion mit dem Benutzer zwar nicht notwendig, es sollte ihm aber zumindest möglich sein, einen solchen Prozess wenigstens in Ansätzen nachzuvollziehen und auf einige Parameter eines solchen Vorgangs bei Bedarf Einfluss nehmen zu können. Eine solche Vorgehensweise bietet den Vorteil, den Benutzer in den Vorgang der Anfrageerweiterung mit einzubeziehen und so einerseits von seinem Wissen zu profitieren und ihm gleichzeitig die Möglichkeit zu geben, dieses Wissen in Beziehung zu setzen zu den verwendeten Strukturen und dem auf diese Weise im System mittels der verwendeten Dokumentationssprache explizierten Wissen. Während hier die Modifikation der Suchanfrage vom System selbst initiiert und vorgenommen wird und der Nutzer möglicherweise lediglich über diese Schritte informiert wird, besteht auf der anderen Seite auch die Möglichkeit, die Entscheidung über eine mögliche Erweiterung oder andersartige Modifikation der Suchanfrage sowie deren Gestaltung dem Benutzer zu überlassen.

4.2.3 Expansion („Query-expansion“)

Einen Sonderfall der Modifikation einer ersten Suchanfrage stellt die Ausweitung dieser Suchanfrage nach bestimmten Vorgaben dar. So können beispielsweise Anfragen, die nicht die gewünschte Treffermenge oder sogar Nulltreffermengen ergeben haben, durch die Hinzunahme spezifischerer Unterbegriffe des ursprünglichen Suchterms verändert werden und für erneute Anfragen an das System genutzt werden. Eine solche Expansion kann bereits mit einfachen, hierarchisch strukturierten Vokabularen wie herkömmlichen Thesauri vorgenommen werden. Die Expansion einer Suchanfrage durch zusätzliche Suchterme hat eine Erhöhung des „*Recalls*“ zur Folge, die allerdings mit einer Einbuße bei der „*Precision*“ einhergehen kann.⁴⁵ Um die Erweiterung der Suchanfrage präzise gestalten und somit die Veränderung der Treffermenge möglichst genau steuern zu können, sollte versucht werden, die Expansion sinnvoll zu begrenzen um ein zu unspezifisches Ausufern der Suche zu verhindern. Eine Möglichkeit die Ausweitung der Suche zu begrenzen und somit eine unkontrollierte Explosion der Ergebnismenge zu verhindern, liegt in der Festlegung der Anzahl an Hierarchieebenen die bei der Erweiterung berücksichtigt werden sollen.

Dies ist vor allem bei umfangreichen und durch zahlreiche Hierarchieebenen sehr kleinschrittig strukturierten Vokabularen sinnvoll.

Die Möglichkeiten der gezielten Gestaltung und Kontrolle einer solchen Erweiterung einer Suchanfrage hängen unmittelbar von den Strukturen vor allem von der Qualität der Relationierung der verwendeten Dokumentationssprache ab. So können anhand entsprechender Strukturen auch andere Arten der Modifikation und Expansion denkbar. So könnten beispielsweise

⁴⁵ Zum Verhalten von Recall und Precision bei relationsgestützten Anfrageerweiterungen vgl. Efthimiadis, E. N. Interactive query expansion: A user-based evaluation in a relevance feedback environment. Journal of the American Society for Information Science 51, (2000) 989-1003

nur Terme aus bestimmten Klassen oder Begriffe die mit dem ursprünglichen Suchbegriff über bestimmte Relationstypen verbunden sind in die semantische Umfeldsuche einbezogen werden. Voraussetzung für solche Methoden sind die differenzierte Gestaltung der hinterlegten Vokabularien sowie die Konzeption entsprechender Retrievalkomponenten. Die Möglichkeiten, welche sich durch eine detaillierte und differenzierte Relationierung des Vokabulars für die Orientierung im begrifflichen Umfeld ursprünglicher Suchterme sowie für die Modifikation und Expansion von Suchanfragen ergeben, sollen im fünften Kapitel ausführlicher dargestellt werden⁴⁶.

⁴⁶ Zu den Möglichkeiten von Erweiterungen von Suchanfragen durch die Verwendung kollektionsunabhängiger Wissenstrukturen s.a.

Efthimiadis, E. N. Query expansion. *Annual Review of Information Science and Technology* 31 (1996) 121-187 sowie

Efthimiadis, E. N. Interactive query expansion: A user-based evaluation in a relevance feedback environment. *Journal of the American Society for Information Science* 51, (2000) 989-1003

5 TYPOLOGIE SEMANTISCHER NETZE

Das Problem mangelnder oder zumindest eingeschränkter Interpretierbarkeit mancher semantischer Netze wurde bereits angesprochen. Dies muss allerdings nicht immer als Versäumnis bei der Modellierung interpretiert und als Einschränkung bedauert werden.

Inwieweit die möglichst vollständige Modellierung semantischer Informationen überhaupt möglich ist und welche Elemente auf welche Art in ihren semantischen Eigenschaften definiert werden können, hängt von der Art des jeweiligen semantischen Netzes und der ihm zugrunde liegenden Ontologie ab.

Die Unterscheidung verschiedener Typen semantischer Netze kann nach technischen, logischen oder auch gestalterischen Aspekten vorgenommen werden. Alle diese Aspekte müssen bei der Analyse und Beurteilung vorhandener Netze, aber auch bei der Konzeption und Modellierung neuer Ontologien berücksichtigt werden.

Bezüglich der logischen Eigenschaften der Ontologien sowie der sie konstituierenden Komponenten kann man ebenfalls verschiedene Typen unterscheiden. Eine im Zusammenhang mit der Weiterentwicklung von Dokumentationssprachen wichtige Unterscheidung ist die zwischen axiomatischen und terminologischen Ontologien wie sie von Fischer⁴⁷ vorgenommen wird und wie sie zur Verdeutlichung von unterschiedlichen Gestaltungsmöglichkeiten sowie der Grenzen der logischen Modellierung in dieser Arbeit übernommen werden soll: Unter axiomatischen Ontologien versteht man danach Konstrukte, bei denen alle zu deren Modellierung verwendeten Komponenten, also alle Objekte und Relationen aus einzelnen Elementarklassen oder semantischen Primitiven nach strengen logischen Regeln gebildet werden, und welche daher idealerweise auch vollständig logisch interpretierbar sind. Diese Ontologien lassen sich bis auf die ihnen zugrunde liegenden Axiome zurückführen und lassen zumindest theoretisch eine vollständige maschinelle Interpretation zu.

Terminologische Ontologien hingegen stellen eine differenzierte und ebenfalls formal strenge Relationierung bereits bestehender Terme dar, welche an sich nicht weiter in semantische Primitive zerlegt werden können und daher auch nicht für eine entsprechende maschinell vorgenommene Interpretation zur Verfügung stehen. In diesen terminologischen Ontologien können die verwendeten Begriffe selbst nur intellektuell interpretiert werden. Eine Modellierung logischer Zusammenhänge ist in diesen Ontologien nur über die Relationierung zu erreichen. Wo genau die Grenzlinien zwischen Terminologien und kontrollierten Vokabularen einerseits und Ontolo-

⁴⁷ Fischer, Dietrich H. From Thesauri towards ontologies? In: *Advances of Knowledge Organization*, vol. 6 (1998) S. 18-30 hier vor allem die Definitionen S.18

gien andererseits verlaufen ist nicht präzise zu bestimmen. Die von Reimer vorgenommene Unterscheidung zwischen Terminologien und Ontologien betont zwar den „formalen Charakter der Begriffsdefinition“ in einer Ontologie, aber auch die Tatsache, dass „Begriffe in einer Ontologie durch [...] Beziehungen zu anderen Begriffen näher charakterisiert [werden.]“⁴⁸

Diese Unterscheidung zeigt sowohl die Verwandtschaft beider Begriffe und unterstreicht auch das Entwicklungspotenzial von Terminologien und kontrollierten Vokabularen hin zu differenzierten Ontologien.

Diese Unterscheidung soll hier ausdrücklich betont werden, um die Eigenarten und Einschränkungen terminologischer Ontologien aufzuzeigen, aber vor allem auch deren Grenzen bezüglich einer weitergehenden maschinellen Interpretation, wie diese bei axiomatischen Ontologien möglich zu sein scheint, zu betonen.

So soll es im folgenden Kapitel auch nicht darum gehen, anhand aller aufgezeigten Gestaltungsmöglichkeiten eine Dokumentationssprache in Form eines semantischen Netzes neu zu konzipieren und von Grund auf aufzubauen. Vielmehr soll gezeigt werden, wie durch die selektive Anwendung einiger spezielle Gestaltungsprinzipien die Strukturiertheit und Expressivität einer bestehenden und bereits zur Indexierung umfangreicher Bestände gebrauchten Dokumentationssprache verbessert werden kann, und so ihre Eignung für Verfahren des Information Retrieval in elektronischen Umgebungen ebenfalls verbessert werden kann.

⁴⁸ Reimer, Ulrich. Wissensbasierte Verfahren der Organisation und Vermittlung von Information. In: Kuhlen, Rainer; Thomas Seeger; Dietmar Strauch (Hrsg.) Grundlagen der praktischen Information und Dokumentation. 5., völlig neu gefasste Ausgabe. Bd. 1 Handbuch zur Einführung in die Informationswissenschaft und – praxis München: Saur, 2004. S. 155-166 hier besonders S. 157

6 EXEMPLARISCHE RELATIONIERUNG EINES TEILGEBIETS DER SWD

In diesem Kapitel soll die Übertragung der in vorangegangenen Kapiteln dargestellten Überlegungen und Möglichkeiten auf ein bereits bestehendes Vokabular am konkreten Beispiel dargestellt werden. Dies soll durch eine differenzierte Relationierung des Gegenstandsbereichs „Theater“ aus der Sachgruppe 15.1 der Schlagwortnormdatei (SWD) geleistet werden. Sowohl die Konzeption und der Aufbau eines Relationeninventars als auch die Modellierung der semantischen Zusammenhänge innerhalb des Gegenstandsbereichs anhand dieses Inventars sollen im Anschluss entwickelt und folgendermaßen dargestellt werden: Nach einer ersten Analyse der Situation sowie Anmerkungen zu Struktur und Umfang der SWD in ihrer heutigen Form soll die Entwicklung des Relationeninventars aus dem Gegenstandsbereichs sowie dessen Strukturierung dargestellt werden. Hierbei bietet der Gegenstandsbereich „Theater“ zahlreiche Möglichkeiten die Eingrenzung, Herausbildung und spätere präzise Modellierung einzelner Relationen sowie die Gestaltung einer hierarchischen Taxonomie modellhaft zu entwickeln.

Die Schlagwortnormdatei versteht sich als ein kooperativ geführtes Werkzeug zur inhaltlichen Erschließung. Zusammen mit den *„Regeln für den Schlagwortkatalog“* (RSWK) bildet es die Grundlage für eine zumindest ihrem Anspruch nach syntaktische Methode der Indexierung, sowie die dementsprechend für die Formulierung von Rechercheanfragen zu verwendende Dokumentationssprache. Ursprünglich entstanden aus einer Schlagwortliste und konzipiert zur Erschließung von Beständen und deren Nachweis in einem konventionellen Schlagwortkatalog besitzt die SWD ein sehr heterogenes, vielgestaltiges und äußerst umfangreiches Vokabular.

Im Folgenden sollen diese Besonderheiten dargestellt und wo möglich ihre Entstehung aus der historischen Entwicklung erklärt werden.

Aus einer Orientierung an möglichst engen und idealerweise koextensiven Schlagwörtern sowie aus der Praxis der kooperativen Führung ergibt sich ein für eine Dokumentationssprache ungewöhnlich großer Umfang. Dies stellt zwar für die maschinelle Verarbeitung erst einmal kein Problem dar, erfordert aber besonders gut entwickelte Hilfsmittel für die Prozesse der Orientierung sowie der Navigation. Da die Realisierung dieser Funktionen größtenteils auf der Relationierung des gesamten Vokabulars aufsetzt, kommt dieser gerade in solch umfangreichen Dokumentationssprachen eine besondere Bedeutung zu. Für die Gestaltung und kontinuierliche Pflege und also auch der Erweiterung und Präzisierung einer Dokumentationssprache sind die Definition neuer Terme und deren Integration in die bestehenden Strukturen wichtig. Bezüglich der Ansetzung neuer Deskriptoren setzt die SWD auf das in den *„Regeln für den Schlagwortkatalog“* (RSWK) festge-

legte „Gebräuchlichkeitsprinzip“⁴⁹ welches jeweils für das einzelne Schlagwort angewandt wird, aber nicht das gesamte Vokabular mit seinen vielfältigen Begriffsbeziehungen und Zusammenhängen ganzheitlich gestaltet. Außerdem lässt sich dieses Gebräuchlichkeitsprinzip dem Benutzer nur sehr bedingt vermitteln, weshalb die Schaffung zahlreicher alternativer Bezeichnungen für einen Begriff als alternative Sucheinstiege sowie deren Relationierung notwendig ist. Auch in diesem Bereich ist eine angemessene und daher funktionale Relationierung eine unabdingbare Voraussetzung für ein gelingendes Retrieval.

Die Entwicklung von einfachen, teils alphabetisch, teils thematisch geordneten Schlagwortlisten hin zu größtenteils hierarchisch strukturierten Thesauri über mehrere Jahrzehnte, wie sie die Entstehung und schrittweise Weiterentwicklung der Schlagwortnormdatei darstellt, ist von einer zunehmenden Differenzierung und Strukturierung des verwendeten Vokabulars gekennzeichnet. Der damit einhergehende potenzielle Funktionszuwachs für die Suche ist zwar prinzipiell gegeben, wird allerdings durch einige gravierende methodische Unzulänglichkeiten und eine teilweise unzureichende Umsetzung beeinträchtigt. So ist die Relationierung sowohl was Umfang und Vollständigkeit angeht aber auch in Hinblick auf die Typen der verwendeten Relationen für die Realisation und Unterstützung der oben dargestellten Funktionalitäten ungenügend. Mag dieses Vokabular so wie das seine Anwendung bestimmende Regelwerk für die Verwendung in einem konventionellen Schlagwortkatalog noch ausreichend gewesen sein, so sind diese Werkzeuge für die Verwendung in einer elektronischen Umgebung nicht mehr geeignet. Zahlreiche Gründe für diese mangelnde Eignung stehen in Verbindung mit der „gewachsenen“ Struktur der SWD.

Aufgrund der historischen Entwicklung der verschiedenen Methoden und Praktiken der inhaltlichen Erschließung sowie der daraus resultierende langjährige Anwendung der RSWK, scheint heute noch immer eine starke Konzentration auf eine möglichst detaillierte Erschließung des einzelnen Dokuments mit möglichst präzisen und idealerweise koextensiven Schlagwörtern zu bestehen.

Auch die Indexierungspraxis, die sich dieses Vokabulars bedient, nutzt bei weitem nicht alle möglichen Funktionalitäten und bleibt dem Prinzip des engen Schlagworts und einer Philosophie der Schlagwortkette als autonomem Gesamtkunstwerks verhaftet. Diese Mentalität mag sich aus der dargestellten historischen Entwicklung erklären und ist vielleicht auch aus der Perspektive der Indexierer, die dieses Vokabular als Grundlage und Werkzeug für ihre tägliche Arbeit der intellektuellen Erschließung benutzen, zu verstehen, birgt aber für die Weiterentwicklung und Integration des Vokabulars in neue, elektronische Umgebungen erhebliche Probleme.

Die Beschäftigung mit dem einzelnen Indexat muss immer von Überlegungen zur Funktionalität der geleisteten Erschließung für eine Suche in

⁴⁹ Zum Gebräuchlichkeitsprinzip vgl. Regeln für den Schlagwortkatalog 3. Aufl. Berlin: Deutsches Bibliotheksinstitut, 1998. hier § 9.5

einer elektronischen Umgebung wenn schon nicht geleitet so doch zumindest durch sie ergänzt werden. Gerade in Zeiten, in denen Benutzer aus anderen Bereichen von Anwendungen und Produkten leistungsfähige und intuitiv zu bedienende Suchkomponenten gewohnt sind, ist es unerlässlich, bei der Erschließung bibliographischer Ressourcen auch die Perspektive der Benutzer einzunehmen und deren Erwartungen entsprechende leistungsstarke Recherche- und Orientierungsmöglichkeiten durch weiterentwickelte Retrievalsysteme zu bieten.

Derzeit wird die Recherche vor allem durch die Tatsache erschwert, dass das Vokabular, welches zur Erschließung benutzt wird nicht in einer angemessenen, also übersichtlichen und navigierbaren im Idealfall sogar intuitiv zugänglichen Form vorliegt. Das Prinzip des engen Schlagworts sowie die Regelung bei der Ansetzung von neuen Schlagwörtern dem Prinzip der Gebräuchlichkeit zu folgen erschweren den Zugang zum Vokabular erheblich,⁵⁰ die Navigation innerhalb des Vokabulars wird durch eine ungenügende und zudem teilweise recht lückenhafte Relationierung erschwert.⁵¹ Die Anwendung moderner Suchstrategien wie diese im vierten Kapitel beschrieben wurden ist aufgrund der nur als rudimentär zu bezeichnenden Relationierung nicht möglich oder zumindest nicht sehr Erfolg versprechend.

Ein erheblicher Teil dieser Schwierigkeiten könnte unter anderem durch eine differenziertere Relationierung, eine damit einhergehende semantische Anreicherung, sowie neue Indexierungsregeln behoben werden. Zahlreiche Lücken und Inkonsistenzen, die zwar für einen Benutzer durch eigenes Wissen und intellektuelle Arbeit zu überbrücken oder zu korrigieren sind, stellen für den Einsatz dieses Vokabulars in Information Retrieval Systemen beträchtliche Probleme dar. Die für die Verwendung in elektronischen Umgebungen oder den Einsatz semantischer Technologien geforderte formale Strenge und strukturelle Konsistenz sind nicht in ausreichendem Maße vorhanden. Eine sinnvolle Umsetzung des Vokabulars in einem System, das den Benutzer bei der Formulierung, sowie der Modifikation oder Expansion seiner Suchanfrage unterstützt und ihm zu diesem Zweck eine umfassende Orientierung im verwendeten Erschließungsvokabular ermöglicht, ist mit einem undifferenziert strukturierten Vokabular nur sehr bedingt zu leisten. Die Integration einer in ihren relationalen Strukturen nur mäßig entwickelten Dokumentationssprache ohne deren gründliche intellektuelle Überarbeitung und grundlegenden Neustrukturierung erscheint wenig sinnvoll, da die auf ihr aufsetzenden Systeme immer hinter den heute möglichen Retrievalfunktionalitäten und somit letzten Endes auch hinter den Erwartungen der Benutzer zurückbleiben müssen.

⁵⁰ Vgl. Vgl. Hubrich, Jessica. Input und Output der Schlagwortnormdatei (SWD). Aufwand zur Sicherstellung der Qualität und Möglichkeiten des Nutzens im OPAC. Master Thesis. (2005) (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft Band 49)

⁵¹ Semantische Anreicherung der Schlagwortnormdatei. Winfried Gödert [u. a.] Ergebnisbericht Stand Juni 2004. Fachhochschule. Institut für Informationswissenschaft 2004 (unveröffentlicht)

Dies alles lässt Ansätze zu einer Verbesserung der Relationierung und einer semantischen Anreicherung des verwendeten Vokabulars unter Zuhilfenahme semantischer Technologien wünschenswert und sinnvoll erscheinen, und auch den antizipierten Aufwand gerechtfertigt erscheinen.

Die nachträgliche Erweiterung oder Veränderung einer bestehenden Relationierung ist immer mit einem enormen Aufwand verbunden. Dies gilt insbesondere dann, wenn neue, bislang nicht verwendete Relationstypen in die Modellierungsmöglichkeiten integriert werden und dadurch gewohnte Relationierungspraktiken nicht mehr angewandt werden können, oder zumindest modifiziert werden müssen. Selbst wenn man davon ausgeht, dass ein voll ausgearbeitetes, übersichtlich strukturiertes und daher gut zu handhabendes Relationeninventar nebst entsprechender Verwendungsregeln vorliegt, ist der zu leistende Aufwand der zur Umsetzung dieser Regeln innerhalb eines bestehenden Vokabulars benötigt wird zugegebenermaßen enorm.

Betrachtet man nur den Aufwand, der alleine zur nachträglichen Vervollständigung lückenhafter, größtenteils monohierarchischer Hierarchieleitern⁵² in Begriffsgefügen wie der SWD benötigt wird, wird deutlich, dass die Ausweisung zahlreicher neuer, differenzierter Relationen anhand eines erweiterten und komplexer gewordenen Inventars mit einem beträchtlichen Arbeitsaufwand verbunden ist.

Wird die Arbeit der Relationierung allerdings in Zukunft gleich bei der Neuansetzung von Schlagwörtern als integraler Bestandteil des Ansetzungsprozesses verstanden und erledigt, ist die vom einzelnen Indexierer zu leistende Arbeit zwar immer noch anspruchsvoll und umfangreich aber arbeitsökonomisch durchaus zu vertreten.

Bedenkt man weiter, dass sich die beschriebenen Arbeiten auf die Dokumentationssprache also die Grundlage einer effizienten und funktionalen Erschließung beschränken und eine neuerliche Erschließung bereits indexierter Bestände nicht zwingend notwendig ist, relativiert sich der Aufwand. Wie in anderen Bereichen steht die zu leistende intellektuelle Arbeit in einem proportionalen Verhältnis zu dem dadurch ermöglichten Zuwachs an Funktionalität und Effizienz.

6.1 Material / Ausgangslage

Als Gegenstandsbereich anhand dessen die Relationierung eines Vokabulars entwickelt und erarbeitet wurde und illustriert werden soll, wurde der Bereich „Tanz/Theater“ mit der Systemstelle 15.1 in der SWD gewählt. Dieser Bereich war bereits im Projekt „Multilingual Access to Subjects“ (MACS)

⁵² Vgl. Hubrich, Jessica. Input und Output der Schlagwortnormdatei (SWD). Aufwand zur Sicherstellung der Qualität und Möglichkeiten des Nutzens im OPAC. Master Thesis. (2005) (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft Band 49)

⁵³ zur Erarbeitung von Schlagwortkonkordanzen bearbeitet worden. Im Rahmen dieses Projekts waren für verschiedene Gegenstandsbereiche Konkordanzlisten mit den begrifflichen Entsprechungen aus drei international verwendeten Referenz-Vokabularen erstellt worden, um auf diese Weise einen multilingualen und idealerweise sprachunabhängigen Zugang zu unterschiedlich erschlossenen Beständen zu ermöglichen. Hierzu war aus der Sachgruppe „Tanz/Theater“ ein Kernbereich von 267 Einzelbegriffen herausgearbeitet worden, welcher das verwendete Vokabular enthielt und den Gegenstandsbereich in konzentrierter Form darstellte. Das Termmaterial dieser Liste grenzte den Gegenstandsbereich noch einmal ein und wurde zur begrifflichen Ausgangslage.

Darüber hinaus schien der gewählte Gegenstandsbereich für das Vorhaben einer Konzeption eines Relationeninventars geeignet, da sich früh abzeichnete, dass zwischen vielen der darin enthaltenen Begriffen eine Vielzahl verschiedener Relationen finden ließ. Das Vorkommen unterschiedlicher Begriffstypen wie Berufs- Gattungs- und Formbezeichnungen sowie eine Vielzahl von Theater- Bühnen- und Aufführungsformen waren gute Voraussetzung für eine umfassende Analyse des Gegenstandsbereichs sowie die Entwicklung eines differenzierten Relationenfundus. Darüber hinaus sollten sich die Möglichkeit einer deutlichen Unterteilung der Begriffstypen für den Prozess der Typisierung und der Reduktion der ursprünglichen Relationenvielfalt als nützlich erweisen.

Die unzureichende Relationierung soll am zentralen Topterm „Theater“ demonstriert werden. Wird bei einem Einstieg mit diesem, das gesamte Themengebiet subsumierende Deskriptor die semantische Umgebung über die Auswahl der nicht weiter spezifizierten Begriffsrelation „Unterbegriff“ (UB) erweitert, so zeigt sich ein eher unübersichtliches Feld unterschiedlichster Deskriptoren. Neben der großen Menge ist es vor allem die Verschiedenartigkeit der Begriffe, die alle mit dem gleichen Relationstyp mit dem Oberbegriff „Theater“ verbunden sind, die eine schnelle und thematisch geleitete Orientierung erschwert. (Abb. 3 und 4) Dazu muss gesagt werden, dass sich diese Situation zum Teil aus der Praxis der Erstellung und Pflege sowie der Datenhaltung und Darstellung der SWD ergibt: In den einzelnen Datensätzen werden jeweils nur die dem jeweiligen Begriff zugeordneten Oberbegriff oder auch mehrere Oberbegriffe ausgewiesen, nicht aber die inverse Relationierung mit den Unterbegriffen. Daher ist eine solche Übersicht über alle mit einem Deskriptor verbundenen Begriffe, vor allem die Menge der ihm zugeordneten Unterbegriffe aus der isolierten Betrachtung der einzelnen Datensätzen nicht zu gewinnen. Die vollständigen Zusammenhänge, die über alle Relationstypen hergestellt werden, sind in ihrer Gesamtheit nicht dargestellt. Stünde eine solche Ansicht wie sie in Abb. 3 für die einzelnen Deskriptoren, sowie die entsprechenden Baumdarstellungen der hierarchischen Begriffsleitern (Abb. 4) bereit, könnten derartige

⁵³ Vgl. Homepage des MACS Projekts <https://macs.vub.ac.be/pub/> sowie die Termliste unter [https://macs.vub.ac.be/pub/system/files?file=16trilingual_from_swd_\(theater\).pdf](https://macs.vub.ac.be/pub/system/files?file=16trilingual_from_swd_(theater).pdf)

Schwierigkeiten und Inkonsistenzen leichter erkannt und vermieden werden. Im Falle des Deskriptors „Theater“ sind sowohl der Umfang der Liste, aber vor allem die Heterogenität der Begriffe deutlich als Problem zu erkennen. Die Verschiedenartigkeit der Aspekte, nach welchen die Unterbegriffe jeweils ausgewählt wurden ist deutlich zu erkennen, ist aber formal nicht dokumentiert. Diese Aspektwechsel, welche bereits problematisch sind, wenn sie über mehrere Ebenen einer Hierarchieleiter verteilt vorkommen und die hier im unmittelbaren Begriffsumfeld eines einzelnen Deskriptors auf ein und derselben Hierarchieebene auftauchen, bergen für die Verwendung solcher begrifflicher Strukturen in elektronischen Umgebungen und Information Retrievalsystemen mit ausgefeilten Suchkomponenten erhebliche Probleme: Die unterschiedlichen Aspekte, welche für die Formulierung und Präzisierung einer Anfrage durchaus Relevanz besitzen, können aufgrund der formal unzureichenden semantischen Auszeichnung nicht präzise angesprochen werden und daher nicht sinnvoll für die Navigation oder die Formulierung einer Suchanfrage genutzt werden. Aber auch für den kognitiv begabten Benutzer ist eine solche undifferenzierte Darstellung nicht unproblematisch. Die Verwendung ein und desselben Relationstyps (UB) zur Anbindung dermaßen zahlreicher und unterschiedlicher Begriffe, suggeriert, dass es sich dabei tatsächlich um dieselbe Art von begrifflicher Beziehung mit den gleichen logischen und semantischen Eigenschaften handelt. Die Unterschiedlichkeit der Begriffe zeigt jedoch, dass dies nicht der Fall ist. So sind beispielsweise die Aspekte, nach welchen die Begriffsbeziehungen zwischen dem Deskriptor „Theater“ und den Unterbegriffen „Musiktheater“ und „Staatstheater“ geordnet sind, deutlich von einander zu unterscheiden, ohne dass dies jedoch durch unterschiedliche Relationstypen oder zumindest durch eine Spezifizierung der einen verwendeten Relation expliziert worden wäre. Diese Verschiedenheit der Aspekte kann vom Benutzer zwar anhand seines eigenen impliziten Weltwissens erkannt, interpretiert und verstanden werden, die gezielte Nutzung solcher Feinheiten für die Navigation und das Retrieval sind allerdings nicht möglich. Daher wäre beispielsweise eine Eingrenzung der Unterbegriffe nach bestimmten Aspekten, wie sie zur sinnvollen Kontrolle einer Anfrageerweiterung benutzt werden könnte, nach einem bestimmten Aspekt nicht möglich. Die Erweiterung einer Suchanfrage nach dem Suchbegriff „Theater“ um spezifischere Begriffe, welche beispielsweise nach dem Aspekt der Form, des Genres oder der Funktion unterschieden werden könnten ist anhand einer solchen Relationierung nicht möglich.



Abbildung 3: Topterm „Theater“ in Listendarstellung (links)

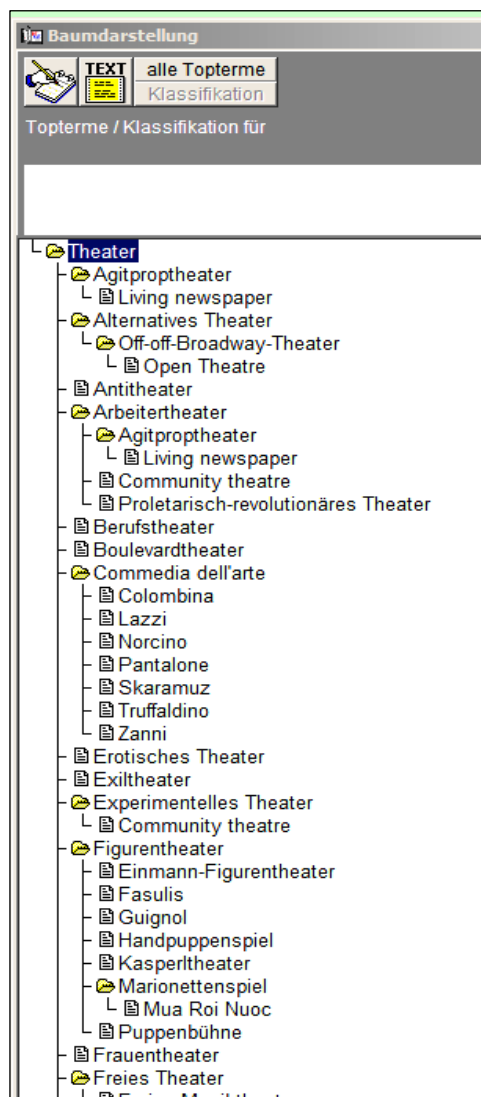


Abbildung 4: Topterm „Theater“ in Baumstruktur (rechts)

Sicher ist der Topterm „Theater“ aufgrund seiner Allgemeinheit und der damit verbundenen Mehrdeutigkeit ein Sonderfall, was die Menge der mit ihm verknüpften Unterbegriffe angeht. Generell lässt sich dieses Phänomen aber auch an anderen zentralen, weniger umfangreich relationierten Begriffen beobachten. Daher ist die Differenzierung der verschiedenen Begriffsrelationen zu einem fein abgestimmten und genau definierten Relationeninventar zur Verbesserung der Retrievalmöglichkeiten wünschenswert.⁵⁴

⁵⁴ Die am Term “Theater” durchgeführte mögliche exemplarische Differenzierung der Relationen ist Tabelle 1 im Anhang dargestellt.

6.2 Werkzeuge / Software

Um die bereits weiter oben dargestellten Modellierungsmöglichkeiten hoch entwickelter Beschreibungsmodelle und Modellierungssprachen für die Zwecke einer differenzierten Gestaltung des Vokabulars nutzen zu können, wird ein Werkzeug benötigt welches die beschriebenen Modelle und Sprachen unterstützt, und so eine ausdrucksstarke Modellierung begrifflicher Strukturen ermöglicht.

Von den zahlreichen existierenden Programmen die zur Konzeption, Modellierung und Kodierung von semantischen Netzen angeboten werden, wurde das von der Stanford-University entwickelte System „Protégé“⁵⁵ verwendet. Neben der guten Handhabbarkeit und Transparenz der Funktionalitäten waren vor allem die Möglichkeiten der Visualisierung sowie die damit verbundene Navigation innerhalb der erstellten Strukturen ausschlaggebend für diese Wahl.

Protégé ist ein Editor-Programm zur Konzeption, Modellierung und Kodierung von Ontologien für unterschiedliche Anwendungen. Je nach Bedarf kann über eine Voreinstellung die benötigte Sprachvariante der „Web Ontology Language“ (OWL) und damit die Palette der Gestaltungsmöglichkeiten ausgewählt werden. Für die Modellierung der Beispielontologie wurde OWL Full gewählt um bei der Gestaltung des Relationeninventars mithilfe der Eigenschaften eine größtmögliche gestalterische Freiheit zu haben.

Das Beschreibungsmodell RDF sowie die darauf aufbauende Modellierungssprache OWL erlauben die Definition von Klassen (*Classes*), Instanzen (*Individuals*) und Eigenschaften (*Properties*). Diese Eigenschaften können in ihren logischen Eigenschaften sowie in den ihnen zuzuweisenden Klassen definiert werden. Durch die unter RDF/OWL mögliche Zuweisung jeweils über- oder untergeordneter Eigenschaften (*Superproperties* bzw. *Subproperties*) ist eine hierarchische Strukturierung des gesamten Relationeninventars möglich. Die Vererbung logischer Eigenschaften zwischen hierarchisch einander untergeordneten Relationstypen erlaubt dabei die Differenzierung der einzelnen Relationstypen entlang der einzelnen Hierarchieleitern. Diese Funktion ist für die Realisierung des hier gemachten Vorschlags von zentraler Bedeutung. Das Relationeninventar ist das semantische Herzstück dieser terminologischen Ontologie

Die Definition verschiedener Klassen ist zwar möglich wurde aber aus bereits dargestellten Gründen nicht zur logischen und strukturellen Modellierung sondern lediglich aus arbeitsökonomischen Gründen verwendet.

Die als Eigenschaften definierten Relationen können jeweils über eine Editor-Funktion für die einzelnen als Instanzen angelegten Begriffe ausgewiesen werden. Die so entstandenen „Triple“ werden zu jeder Instanz übersichtlich dargestellt und können anhand der zugeschalteten Visualisierungskomponente „Jambalaya“ in verschiedenen Formen dargestellt wer-

⁵⁵ Homepage “The Protégé Ontology Editor and Knowledge Acquisition System” - <http://protege.stanford.edu/>

den. Die Auswahl bestimmter Relationstypen sowie die Begrenzung des semantischen Umfelds über die Anzahl der Stufen welche bei der Darstellung der semantischen Umgebung (*Neighbourhood*) berücksichtigt werden sollen, können über entsprechende Filter vorgenommen werden.

Es wurde bereits darauf hingewiesen, dass die hier dargestellte Relationierung sich in vielen Punkten von der Konzeption und Modellierung eines semantischen Netzes unterscheidet. Daher stellt auch der Gebrauch eines Werkzeugs, wie *Protégé*, welches präzise auf solche Beschreibungsmodelle und Modellierungssprachen und deren Funktionalitäten ausgerichtet ist, zur Erstellung solcher terminologisch basierter semantischer Netze einen Kompromiss dar.

Da es sich bei dem erarbeiteten Vorschlag nicht um die Modellierung eines semantischen Netzes von Grund auf handelt, sondern um die nachträgliche Relationierung eines schon bestehenden Begriffsinventars, wurden nicht alle Funktionalitäten und Gestaltungsmöglichkeiten, welche das Programm bietet, genutzt. Im Vordergrund standen die Funktionen zur Auszeichnung und Definition von Relationen, welche zwischen jeweils zwei oder mehreren Begriffen modelliert werden können. Andere Funktionen sollten nicht angewandt werden. So war die Definition und der synthetische Aufbau der einzelnen, teils sehr komplexen Begriffe aus Elementarbegriffen oder einzelnen semantischen Komponenten anhand verschiedener logischer Bedingungen nicht möglich oder sinnvoll. Auf die Konsequenzen eines bereits bestehenden natürlichsprachlich basierten Begriffsinventars wurde bereits an anderer Stelle eingegangen.⁵⁶

Da die einzelnen Begriffe nicht als Elemente von Klassen mit bestimmten als notwendig oder hinreichend definierten Eigenschaften definiert wurden, sondern als nicht weiter bestimmte, und nicht in ihren logischen Eigenschaften ausgezeichnete Elemente aus der SWD übernommen wurden, können sie als Individuen modelliert werden. Dies mag für die Modellierung klassischer semantischer Netze unüblich erscheinen, bietet jedoch den Vorteil, dass die unterschiedlichen tatsächlich bestehenden Beziehungen genau dargestellt werden können. Definierte man die Begriffe als Klassen, so ließen sich über die Angabe der „*Range*“ sowie der „*Domain*“ lediglich die Gruppe der für diese Klasse möglichen, nicht aber die tatsächlich bestehenden Relationen zu anderen Begriffen darstellen. Die Besetzung der Kategorie der Individuen ist zudem auch unproblematisch, da den Begriffen nicht noch zusätzliche Individuen als Instanzen zugeordnet werden müssen, sondern das so erstellte Vokabular zur Erschließung bibliographischer Bestände in einer anderen Anwendung verwendet werden soll.

⁵⁶ s. Kapitel 4.1 dieser Arbeit.

6.3 Technisches Vorgehen

Im Folgenden sollen die einzelnen Schritte, welche zur Umwandlung, Analyse und Anreicherung sowie zur relationalen Differenzierung der SWD Daten vorgenommen wurden im Einzelnen dargestellt werden. Dabei soll auch erörtert werden, welche dieser Schritte möglicherweise automatisiert oder zumindest rationalisiert werden können um eine mögliche Übertragung der dargestellten Überlegungen und Vorgänge auf umfangreichere Teilbereiche oder sogar das gesamte Vokabular zumindest teilweise möglichst arbeitsökonomisch gestalten zu können.

Ausgangsmaterial für die Relationierung des Gegenstandsbereichs „Theater“ sind die in der SWD vorhandenen Deskriptoren zu diesem Themengebiet, sowie die zwischen ihnen bestehenden Relationen wie sie in der SWD abgebildet sind. Die Deskriptoren wurden über eine Recherche mit Notationstrunkierung ausgewählt und im SWD-spezifischen MAB-Format als Text-Datei exportiert. Für die Sichtung der Gesamtmenge der Deskriptoren und besonders für die genaue Analyse der bereits bestehenden Relationierung ist eine angemessene Darstellungsform mit verschiedenen Ansichtsoptionen und Navigationsmöglichkeiten notwendig. Daher wurden die exportierten Datensätze in einem Zwischenschritt in einer Umgebung dargestellt, die für jeden Datensatz alle bestehenden Relationen übersichtlich und navigierbar darstellt. Die Software *MIDOS Thesaurus* bietet für diese Zwecke eine gut strukturierte, funktionale und transparente Umgebung an. Bereits bei diesem Zwischenschritt während der Bearbeitung zeigte sich die Bedeutung einer übersichtlichen und gut strukturierten Darstellung durchaus vorhandener Strukturen. Die Transformation der SWD-Datensätze in ein mit *MIDOS Thesaurus* kompatibles Format war durch einfache Ersetzungsprozesse, welche auch ohne großen Aufwand automatisiert werden können möglich. Zu beachten ist, dass diese Umwandlung lediglich ein für die Arbeit der Analyse und Konzeption notwendiger Zwischenschritt war, der die Ansicht sowie die Navigationsmöglichkeit bereits vorhandener Strukturen verbessern sollte, und zu diesem Zeitpunkt noch keine strukturellen oder relationalen Veränderungen des Vokabulars vorgenommen wurden.

Zwischenergebnis war eine mth-Datei der *MIDOSTheasurus* Anwendung des Gegenstandsbereichs „Tanz und Theater“ der die in der SWD befindlichen Deskriptoren sowie die ausgewiesene Relationierung in Form eines navigierbaren Thesaurus abbildet. Er enthält 1008 Deskriptoren, 3194 Verweise und weist die Relationstypen OB/UB, VB, sowie Synonyme und Definitionen aus. Die Benutzeroberfläche im reinen Betrachtungsmodus entspricht der Ansicht, welche auch die Mitarbeiter der Deutschen Nationalbibliothek (DNB) seit einiger Zeit als Hilfestellung zur Orientierung und Navigation in der SWD auch im Zusammenhang mit der DDC haben. Die Darstellung des Begriffsraums in einer Thesaurusansicht war lange Zeit keine Selbstverständlichkeit. Die Praxis, die Anlage und Nutzung von Deskriptoren ausschließlich über die Ansicht der einzelnen Datensätze und nicht über

die Zusammenhänge und Beziehungen zwischen den einzelnen Begriffen zu koordinieren kann einige der Inkonsistenzen und Fehler in der Relationierung erklären. Dies zeigt die Notwendigkeit einer angemessenen Darstellung der Relationsstrukturen.

Unter Zuhilfenahme dieses Zwischenprodukts wurden 267 ausgewählte Deskriptoren mitsamt der bestehenden Relationierung in ein Projekt unter der Software *Protégé* übertragen. (Datei: Theater-alt.pprj) Dieser Schritt erfolgte manuell, da zu diesem Zeitpunkt bereits verschiedene Modellierungsmöglichkeiten und andere Funktionalitäten zur Gestaltung des übernommenen Vokabulars erprobt werden sollten, und dabei Vor- und Nachteile verschiedener Ansätze überprüft werden konnten. Grundsätzlich ist aber auch diese Übertragung aufgrund der genau definierten Vorgaben für das Zielformat RDF/OWL automatisch möglich. Dabei ist die Umwandlung von einem linearen, datensatzorientierten hin zu einem hochgradig komplexen und vielfach ineinander verschachtelten Skriptformat durch entsprechende Ersetzungs- und Umstrukturierungsprozesse zu leisten.

Die in der vom MACS-Projekt übernommenen Auswahlliste enthaltenen Deskriptoren stellen den Kernbereich des Gegenstandsbereichs dar. Dieser Kernbereich enthält eine begrenzte Menge an Deskriptoren der SWD sowie die zwischen ihnen bestehende undifferenzierte Relationierung. Er lag nun unter „Protégé“ in einer Umgebung vor, die zahlreiche Funktionalitäten bot, um eine differenzierte Modellierung eines komplexen Relationengeflechts zu unterstützen und das Ergebnis transparent darzustellen und für das Information Retrieval anzubieten.

Dabei verfolgte die Übertragung der bestehenden Strukturen in die neue Modellierungsumgebung keinesfalls die Absicht, die Relationierung dieses Vokabulars mit seinen inkonsistenten und lückenhaften Beziehungen als Ausgangspunkt für die Bildung des neuen Inventars oder zur Orientierung für die Konzeption der neuen Relationierung zu benutzen. Vielmehr war die durch verschiedene neue Programme ermöglichte Betrachtung der bestehenden Strukturen eine wichtige Voraussetzung für eine gründliche Analyse. Zahlreiche der an anderer Stelle bereits dargestellten Unzulänglichkeiten und Inkonsistenzen wurden durch eine solche neue Darstellung der alten Strukturen noch einmal verdeutlicht, andere überhaupt erst evident. Die Darstellung der bestehenden begrifflichen Struktur in den unterschiedlichen Darstellungsformen der jeweiligen Programme verdeutlicht auch noch einmal die Relevanz und das Potential, einer differenzierten und aussagekräftigen Relationierung für verschiedene Anwendungsbereiche.

Für den Prozess der Findung, Analyse und Inventarisierung der neuen, auszuweisenden Relationen, wurden aus dem oben beschriebenen Zustand die Deskriptoren ohne die bestehenden Relationen in ein zweites „Projekt“ übernommen. Diese eingegrenzte und unrelationierte Menge an Deskriptoren bildete somit den Ausgangspunkt für die Überlegungen zur Erstellung eines Relationeninventars wie dies in Kapitel 5.6 dargestellt werden wird. In einem nächsten Schritt wurden dann die Gesamtheit der Deskriptoren an-

hand der im neu erarbeiteten Inventar enthaltenen und definierten Relationen in der Umgebung von *Protégé* modelliert, und die daraus entstehenden Strukturen in der entsprechenden Form kodiert. Somit steht der neu relationierte Kernbereich des Gegenstandsbereichs „*Theater*“ für verschiedene Funktionen der Visualisierung sowie des Retrievals zur Verfügung.

Die technische Umsetzung, insbesondere die Kodierung sowie die zur Nutzung des Vokabulars erstellten Anwendungen haben noch stärker als die theoretischen Überlegungen ausschließlich experimentellen Charakter. Solche Komponenten können aber durchaus als separate Produkte unabhängig und nachgeordnet erstellt werden. Auf die Möglichkeiten und Anforderungen an derartige Anwendungen wird noch zu einem späteren Zeitpunkt im sechsten Kapitel eingegangen werden. Im Rahmen dieser Arbeit dienen diese Anwendungen lediglich dazu, bestimmte Prinzipien und Wirkmechanismen zu illustrieren und an einem konkreten Gegenstandsbereich entwickeln zu können. Insbesondere die Möglichkeiten der Visualisierung, die von dem Programm sowie zusätzlich verwendeten Plug-ins⁵⁷ bereitgestellt wurden, sind für die dargestellten Zwecke nur bedingt geeignet, und konnten nicht oder nur ansatzweise an die Erfordernisse und Besonderheiten dieser Modellierung angepasst werden. Gleiches gilt für die Möglichkeiten der Navigation und Recherche. Hier konnte nur auf einige grundlegende Funktionalitäten, welche schon in der verwendeten Software implementiert waren, zurückgegriffen werden. Da auch die Konzeption und Programmierung einer den modellierten Datenstrukturen sowie den der Aufgabenstellung angemessenen Rechercheumgebung den Rahmen dieser Arbeit gesprengt hätte, können hier bestimmte Überlegungen und Vorgehensweisen lediglich theoretisch dargestellt werden. Die von der Visualisierungskomponente angebotenen Query-Funktionen erlauben zwar eine Simulation bestimmter Suchanfragen sowie die Einschränkung und Spezifizierung einzelner gewünschter Relationen, ließen sich aber bezüglich ihres Bedienkomforts sowie der Formulierungsmöglichkeiten für Anfragen an das System noch wesentlich verbessern. Die Realisierung verschiedener Komponenten sowie zusätzlicher Produkte, welche sich aus diesen Überlegungen ableiten lassen, bedarf noch einigen technischen und sicher auch intellektuellen Aufwands um die gewonnenen Erkenntnisse umzusetzen und in konkreten Produkten, technisch ausgereift zu implementieren.

6.4 Bisherige Entwicklungen

Neben zahlreichen Kommentaren zu Weiterentwicklungen und Neuauflagen des Regelwerks sowie zu Fragen der Pflege und Erweiterung des Vokabu-

⁵⁷ Die verwendete Visualisierungskomponente „Jambalaya“ stellt zwar nur eine sehr einfache und kaum an besondere Erfordernisse anzupassende Umsetzung der kodierten Modellierung zur Verfügung, ermöglicht aber eine sehr direkte und auch in ihrer Transparenz und Flexibilität für den Prozess der Konzeption und Modellierung gut geeignete Umsetzung in eine dynamische und interaktiv zu navigierende Visualisierung.

lars⁵⁸ standen immer wieder Fragen nach der grundsätzlichen Eignung der angewandten Methode und der verwendeten Werkzeuge. Zunehmend geriet hier auch die Perspektive der Endnutzer in den Fokus der Überlegungen.

Hauptkritikpunkt waren die geringe Eignung der entstehenden Indexate für maschinelle Verfahren des Information Retrievals sowie die damit verbundene geringe Nachvollziehbarkeit der Verwendung des Vokabulars für den Benutzer.

Bei den Versuchen diese Situation zu verbessern, wurden zwei unterschiedliche Herangehensweisen in Betracht gezogen. Zum einen gab es Überlegungen, ausgehend von dem bestehenden Vokabular die vorhandenen Relationierung durch die Vervollständigung und Verdichtung des Relationennetzes unter Verwendung des bisherigen Relationeninventars zu verbessern.⁵⁹ Dabei würden allerdings bestehende Inkonsistenzen oder fehlerhafte Relationierungen nicht berücksichtigt und korrigiert. Zudem wären die gewünschte Differenzierung der Relationierung und die damit angestrebte Verbesserung der Gestaltungsmöglichkeiten nicht gegeben. Der benötigte Arbeitsaufwand für eine solche aufwendige intellektuelle Kleinarbeit wäre auch bei der Verwendung des bereits bestehenden und äußerst übersichtlichen Relationeninventars dennoch enorm, zumal sich die Prozesse der Analyse bestehender oder fehlender Relationen sowie deren Ergänzung kaum oder nur sehr bedingt rationalisieren und schon gar nicht automatisieren ließen.

Darüber hinaus wurde auch der Einsatz neuer Relationstypen eines notwendigerweise differenzierter entwickelten Inventars diskutiert. So konstatiert Kunz zu dem Stichwort „*Relationierung*“: „Benutzung semantischer Relationen als Navigationshilfe [...] bedingt Weiterentwicklung und Anpassung hierarchischer und assoziativer Relationen“ und empfiehlt eine „Anpassung bislang RSWK-konform nachgewiesener Relationen an OPAC Ansprüche“⁶⁰

Inventare, die eine überschaubare Anzahl von Relationstypen für die Modellierung und Interpretation eines bestimmten Gegenstandsbereichs im Kontext einer bestimmten Anwendung enthalten, existieren bereits in unterschiedlichen Formen. Hier seien beispielsweise die Auflistungen, Inventare oder Taxonomien für Relationstypen von Aitchison⁶¹ oder Milstead⁶², sowie

⁵⁸ Kunz, Martin. „Zur Weiterentwicklung der SWD“ In: Bibliotheksmanagement – Kulturmanagement. Vorträge und Berichte. 24. Österreichischer Bibliothekartag Congress Innsbruck, 3-7.9. 1996 (Biblos-Schriften Band 168) S.177-196

⁵⁹ Semantische Anreicherung der Schlagwortnormdatei. Winfried Gödert [u. a.] Ergebnisbericht. Fachhochschule. Institut für Informationswissenschaft (unveröffentlicht) 2004.

⁶⁰ Kunz, Martin. „Zur Weiterentwicklung der SWD“ In: Bibliotheksmanagement – Kulturmanagement. Vorträge und Berichte. 24. Österreichischer Bibliothekartag Congress Innsbruck, 3-7.9. 1996 (Biblos-Schriften Band 168) S.177-196 hier S. 177.

⁶¹ Aitchison, Jean; Alan Gilchrist, David Bawden. Thesaurus construction and use: a practical manual. London: Aslib 1997. Zur Diskussion dieser Vorschläge s. a.

die zahlreichen Zusammenstellungen und Übersichten nicht-hierarchischer Relationstypen von Neelameghan⁶³ sowie von seiner Arbeit abgeleitete Sammlungen zu nennen. Ebenso gibt es verschiedene Ansätze umfangreichere Sammlungen von Relationstypen in einer hierarchischen Struktur zu ordnen.⁶⁴ Der Findungsprozess, der zu einem solchen Inventar führt, wird dabei allerdings nicht immer erörtert, was zum Teil auch mit der deduktiven, vielleicht sogar intuitiven Herangehensweise bei dessen Schaffung zu tun hat. Auch wenn diese Inventare auf andere Weise gebildet werden und sich daher nicht dazu eignen, die notwendigen Validierungsprozesse zu verdeutlichen, können sie trotzdem zur Orientierung und als Anschauungsmaterial für Überlegungen zur Inventarbildung dienen. Aktuelle Diskussionen dieser Möglichkeiten sowie Ansätze zur Bildung und weitergehenden Verfeinerung von Relationstypen und Inventaren vor dem Hintergrund semantischer Technologien finden sich auch bei Mazzocchi⁶⁵ sowie bei Peters/Weller⁶⁶

Eine weitere Möglichkeit, die Strukturiertheit des Vokabulars der Dokumentationssprache und die sich daraus ergebenden Retrievalfunktionalitäten zu verbessern, kann in der Verbindung des Vokabulars der SWD mit der differenziert entwickelten und bewährten Klassenstruktur der *Dewey Deci-*

Clarke, Stella G. Dextre. Thesaural Relationships. In: Bean, Carol A. Rebecca Green (Hrsg.): Relationships in the Organisation of Knowledge. Dordrecht: Kluwer Academic Publishers. 2001 (Information Science and Knowledge Management 2). S.37-52

⁶² Vgl.: Milstead, Jessica L. Standards for relationships between subject indexing terms. In: Bean, Carol A. Rebecca Green (Hrsg.): Relationships in the Organisation of Knowledge. Dordrecht: Kluwer Academic Publishers. 2001 (Information Science and Knowledge Management 2). S.53-66

⁶³ Beghtol, C.: The facet concept as a universal principle of subdivision. In: Knowledge organization, information systems and other essays: Professor A. Neelameghan Festschrift. K.S. Raghavan K.N. Prasad. (Hrsg.) New Delhi: Ess Ess Publications 2006. S.41-52.

a

Green, R., C.A. Bean: Aligning systems of relationships.

In: Knowledge organization, information systems and other essays: Professor A. Neelameghan Festschrift. K.S. Raghavan and K.N. Prasad. (Hrsg.) New Delhi: Ess Ess Publications 2006. S.111-128.

Neelameghan, A.; I. K. Ravichandra Rao. Non-hierarchical associative relationships: their types and computerization of links. In: Library science with a slant to documentation and information studies (1976) 13, S. 24-42

Neelameghan, A.; Ranjitra Maistra. Non-hierarchical associative Relationships among concepts: identification and typology. (Part A of the FID/CR Report No 18 of the Committee on Classification Research (FID/CR) International Federation for Documentation (FID) Bangalore: Documentation Research and Training Centre 1978

Maistra, Ranjita. Semi-automatic method of preparing thesaurus for a specific subject-field. Part B of the FID/CR Report No 18 of the Committee on Classification Research (FID/CR) International Federation for Documentation (FID). Bangalore: Documentation Research and Training Centre 1978

⁶⁴ Vgl. hierzu Dee, Michel, Taxonomy of subject relationships, Appendix B zu: American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSSS Subject Analysis Committee. (1997)
<http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/msrscu2.pdf>

⁶⁵ Mazzocchi, Fulvio; Melissa Tiberi, Barbara De Santis, Paolo Plini. Relational semantics in thesauri: some remarks at theoretical and practical levels. In: Knowledge Organization Vol. 34 (2007) 4 S. 197-214

⁶⁶ Peters, Isabella; Katrin Weller. Paradigmatic and syntagmatic relations in knowledge organization Systems. In: Information Wissenschaft und Praxis. 59 (2008) 2 S. 100-107

mal Classification (DDC) gesehen werden. Die Integration der DDC deren Gebrauch für den deutschen Sprachraum durch die im Jahr 2005 abgeschlossene Übersetzung und Anpassung ins Deutsche wesentlich erleichtert, wenn nicht sogar erst ermöglicht worden war, bietet die Möglichkeit, die hoch entwickelte Struktur der DDC auch für die SWD nutzbar zu machen. Im Rahmen des CrissCross Projekts zeigte sich, dass sich die Struktur der DDC immer deutlicher als ein zusätzliches, alternatives Strukturgerüst der SWD herausbildet. Dabei dienen die den Begriffen der SWD durch intellektuelle Analyse zugewiesenen DDC Notationen als Bindglied, mit dessen Hilfe die umfangreichen, differenzierten und bewährten Strukturen der DDC auf das Vokabular der SWD indirekt übertragen werden konnten.

Aus der Beschäftigung mit den beschriebenen Untersuchungen und Überlegungen im Rahmen des CrissCross-Projekts, sowie durch die Kenntnis der mangelhaften Relationierung erwuchs die Überlegung, mit dem Versuch einer erweiterten und differenzierten Relationierung der bestehenden Dokumentationssprache eine weitere Möglichkeit zur Neugestaltung und grundlegenden Transformation der SWD zu versuchen.

Die in dieser Arbeit entwickelten Überlegungen zielen auf eine verbesserte Handhabbarkeit der SWD durch die Erweiterung und Differenzierung des zur Strukturierung des Vokabulars verwendeten Relationeninventars.

Der hier gemachte Vorschlag unterscheidet sich von den dargestellten Ansätzen insofern, da hier das zu schaffende Relationeninventar aus einem konkreten, in Form von natürlichsprachlichen Begriffen vorliegenden Gegenstandsbereich eines bereits bestehenden und genutzten Referenzvokabulars entwickelt werden soll. Die Bindung an ein bereits bestehendes Vokabular und die Vorgabe, das anhand des neu zu konzipierenden Relationeninventars strukturierte Vokabular für das Retrieval durch den Benutzer zu verwenden, erhöhen die Notwendigkeit der Bildung eines gleichzeitig allgemeingültigen wie auch hinreichend spezifischen Relationeninventars sowie eine übersichtliche und funktionale Strukturierung dieses Inventars. Eine solche Strukturierung wird durch die hierarchische Gestaltung des Inventars in Form einer Taxonomie angestrebt. Die Möglichkeiten der logischen Modellierung sollen daher sehr selektiv benutzt werden und können nicht den Anspruch erheben, eine weiterhin notwendige kognitive Interpretation durch die maschinelle Verarbeitung logischer Eigenschaften semantischer Strukturen zu ersetzen.

6.5 Strukturierung des Gegenstandsbereichs

Ausgehend von dem vorgefundenen Ausschnitt des bestehenden Vokabulars ging es in den folgenden Schritten darum, die von den Beschreibungsmodellen und Modellierungssprachen vorgesehenen Gestaltungsmöglichkeiten auf die vorliegende Ausgangsmenge an Termen anzuwenden und dabei die Möglichkeiten aber auch Grenzen der Modellierung zu erproben und somit verschiedene Herangehensweisen und Entwicklungsstrategien zur

Modellierung des Gegenstandsbereichs zu entdecken. Hierbei ging es vor allem darum, die beiden wichtigsten Strategien zur Strukturierung und Ordnung solcher Bereiche, nämlich die Klassenbildung sowie den Ausweis semantischer Relationen auf Ihre Tauglichkeit und Umsetzbarkeit hin zu prüfen

6.5.1 Begriffsdefinition durch Klassenbildung

Die Strukturierung des gesamten Vokabulars (Begriffsmenge) erhebt nicht den Anspruch auf logische oder sogar ontologische Validität. Auf den Unterschied zwischen einer logisch valide aufgebauten und von Grund auf durchkonstruierten axiomatischen Ontologie und einer ordnenden terminologischer Ontologie welche zur Ordnung und Strukturierung eines bestehenden Begriffssystems dient, wurde bereits hingewiesen.⁶⁷

Die konsequente Unterteilung und Ordnung aller Begriffe in verschiedene Klassen ist daher nicht für das Retrieval zu verwenden. Bei einer umfassenden und vollständigen Definition aller logischen Bedingungen aller Begriffe könnten solche Angaben für die Formulierung von Begriffen und entsprechende Suchanfragen im Retrieval genutzt werden. Der Gebrauch logischer Attribute stellt bei terminologischen Ontologien aber immer nur notwendige, nie hinreichende Bedingungen dar.

Da sich die Definition der verschiedenen Klassen, vor allem die logischen Zusammenhänge und die sich daraus ergebenden Vererbbarkeiten der Klassenmerkmale nur sehr begrenzt für die Begriffsdefinition anwenden lassen ist jedoch eine Nutzung für das Retrieval nicht möglich.

Trotzdem bietet der Gebrauch von Klassen zur Ordnung der Begriffsmenge Vorteile bei der Erschaffung, Bündelung und Typisierung der Relationen, welche zwischen den einzelnen Begriffen *aller* Klassen bestehen, und kann bei deren notwendiger Validierung von Nutzen sein. So können beispielsweise die „*Domains*“ und „*Ranges*“ bestimmter Relationstypen eingeschränkt werden. Auf diese Weise können Konsistenz und Validität der Relationierung zwar nicht garantiert, aber doch deren Überprüfung zum Zeitpunkt der Relationierung unterstützt werden. So kann beispielsweise durch die Schaffung einer eigenen Klasse für Synonyme eine versehentliche, fehlerhafte Verbindung zweier „Nicht-Vorzugsbenennungen“ durch eine Äquivalenzrelation vermieden werden.

Es soll allerdings an dieser Stelle noch einmal betont werden, dass jegliche Definition verschiedener Klassen im vorliegenden Fall lediglich als Unterstützung von Überlegungen und Experimenten für den Reduktionsprozess zur Schaffung des Relationeninventars zu verstehen ist. Die Nutzung von Klassenzugehörigkeiten oder anderer logischer Eigenschaften für Zwecke des Information Retrieval ist nicht möglich. Die Modellierung der tatsächlich für Navigation und Retrieval funktionalen Strukturen und Zusammenhänge wird ausschließlich über die Relationierung erfolgen.

⁶⁷ Vgl. Kapitel 4.1 dieser Arbeit

6.5.2 Semantische Relationen

Da die Definition von Klassen, wie dargestellt, zur Strukturierung des Vokabulars nur bedingt geeignet ist, kommt der Konzeption und der Verwendung der semantischen Relationen eine umso bedeutendere Schlüsselrolle zu. Die durch sie ermöglichte Definition und Ausweisung innerhalb der Dokumentationssprache bestehender, *a priori* gültiger, differenzierter Begriffsbeziehungen ist die wichtigste Voraussetzung für die Darstellung wichtiger Bezüge und Zusammenhänge und somit zur Strukturierung des Vokabulars der Dokumentationssprache welche in einer das moderne Information Retrieval unterstützenden Form maschinenlesbar und sogar in Ansätzen maschinell interpretierbar vorliegen. Für die maschinelle Verarbeitung ist im Bezug auf die Struktur des Vokabulars die genaue und formal strenge Definition aller verwendeten Relationen wichtig, wobei dies vor allem den Ausweis ihrer logischen Eigenschaften bedeutet. Die durch die gewünschte und für eine differenzierte und daher erst aussagekräftige und funktionale Relationierung bedingte Vielzahl an verschiedenen Relationstypen bietet für die maschinelle Verarbeitung keine größeren Probleme; da aber sowohl bei der Relationierung neu aufgenommener Begriffe durch Indexierer aber vor allem bei der Recherche und allen ihren Teilschritten durch den Benutzer eine intellektuelle Interpretation sowie eine Auswahl- oder Kombinationstätigkeit notwendig ist, ist die Möglichkeit der Orientierung in einem gut strukturiertem und damit übersichtlich gestaltetem Relationenfundus eine absolute Notwendigkeit.

6.6 Definition und Strukturierung eines Relationeninventars

Sollen das Retrieval und die Navigation, wie dargestellt, auf den durch Relationen erstellten Strukturen aufsetzen, so ist den Schritten die zur Modellierung dieser Strukturen führen, entsprechende Aufmerksamkeit zu widmen. Die Anzahl der verschiedenen Relationstypen muss hinreichend differenziert und gleichzeitig sinnvoll begrenzt und das durch diese Auswahl geschaffene Relationeninventar übersichtlich und funktional gestaltet werden. Dieses Relationeninventar ist von zentraler Bedeutung für die Gestaltung der Indexierungssprache sowie der mit ihrer Hilfe geleisteten Erschließung und schließlich vor allem für das Funktionieren des Retrievals in mit diesen Sprachen erschlossenen Beständen. Während im Bereich klassischer Thesauri die Menge der verwendeten Relationen sehr übersichtlich und wenig differenziert ist, und teilweise durch entsprechende Richtlinien festgelegt oder zumindest vorgeschlagen wird,⁶⁸ existieren solche Vorgaben oder

⁶⁸ Neben den einschlägigen DIN Normen (DIN 1463, DIN 31 623 und DIN 2330) existieren für den Aufbau und die Pflege konventioneller Thesauri zahlreiche Handbücher und Leitfäden, hier seien nur die Standardwerke Wersig, Gernot. Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. München: Saur, 1985 (DGD-Schriftenreihe 8). und

gar „Standardlösungen“ für differenziertere Relationeninventare nicht. Auch wenn immer wieder versucht wird, durch zahlreiche Angleichungs- und Abstraktionsprozesse zu allgemeingültigen und universell einsetzbaren Relationeninventaren zu gelangen, gibt es für diese Herausforderung bislang keine überzeugenden Lösungen.⁶⁹ Verschiedene bereits geleistete Ansätze, wie die von der *Classification Research Group* zu Forschungszwecken bewusst in ihrer maximalen Komplexität konzipierte Taxonomie,⁷⁰ bieten zwar wertvolle Einsichten, wären aber in ihrer Anwendung in der Praxis eher unhandlich und wenig funktional. Zu komplex und multidimensional sind die zu schaffenden Sammlungen, zu speziell Ihre Gegenstandsbereiche und Anwendungsgebiete. Außerdem wäre die Verwendung eines bereits bestehenden und daher zwangsläufig sehr allgemeinen Relationeninventars möglicherweise zwar eine aus arbeitsökonomischer Sicht zu begrüßende Herangehensweise, die aber beträchtliche Einbußen in den gestalterischen Modellierungsmöglichkeiten und somit in der Aussagekraft der Indexierungssprache mit sich bringen würde. Daher ist die Erstellung und Verwendung eines eigens für einen bestimmten Gegenstandsbereich angefertigten Relationeninventars unter Gesichtspunkten der Gestaltbarkeit und Aussagekraft von Vorteil. Die Konzeption und Erstellung eines solchen Relationeninventars ist ein komplexer mehrstufiger Prozess. Er setzt eine intensive Auseinandersetzung mit dem gesamten Gegenstandsbereich voraus, wenn das Relationeninventar domänenspezifisch entwickelt, und aus dem Gegenstandsbereich selbst herausgearbeitet werden soll. Ein solches Vorgehen bietet zwar einige Vorteile, wie dies zu einem späteren Zeitpunkt noch dargestellt werden soll, dies ist allerdings nicht die einzige denkbare Methode.

Diese Überlegungen können und sollen nicht zur Auswahl eines einzigen gangbaren Verfahrens führen, sondern sollen Schwierigkeiten, Möglichkeiten und Grenzen eines solchen Prozesses sowie der verschiedenen Methoden darstellen, sowie ein Bewusstsein für Vor- und Nachteile bestimmter Herangehensweisen schaffen.

Grundsätzlich stellt sich bei der Konzeption und Entwicklung solcher Relationeninventare die Frage, ob bei der Findung und Definition der Relationstypen induktiv oder deduktiv vorgegangen werden soll. Im Folgenden sollen diese beiden Methoden dargestellt und diskutiert werden. Dabei sind die Begriffe „induktiv“ und „deduktiv“ nicht als logische oder philosophische Konzepte zu verstehen, sondern sollen lediglich zur Bezeichnung eines Vorgehens und schrittweisen Voranschreitens vom Speziellen zum Allge-

Aitchison, Jean; Alan Gilchrist, David Bawden. Thesaurus construction and use: a practical manual. London: Aslib 1997. genannt.

⁶⁹ Mazzocchi, Fulvio; Melissa Tiberi, Barbara De Santis, Paolo Plini. Relational semantics in thesauri: some remarks at theoretical and practical levels. In: Knowledge Organization Vol. 34 (2007) 4 S. 211

⁷⁰ Dee, Michel, Taxonomy of subject relationships, Appendix B zu: American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSS Subject Analysis Committee. (1997)
<http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/msrscu2.pdf>

meinen und einer entsprechenden Entwicklung spezifischer zu allgemeinen oder allgemeineren Relationstypen und umgekehrt verwendet werden.

Bei einem deduktiven Vorgehen müssten die Relationen aus allgemeinen Kategorien abgeleitet werden, welche durch philosophische oder ontologische Überlegungen gefunden und begründet werden müssten. Dabei böte sich auch die Möglichkeit, ausgehend von einigen wenigen bereits vorhandene oder allgemein bekannte Relationstypen, durch eine schrittweise Differenzierung spezifischer zu gestalten und so weitere, spezifischere Relationstypen zu erhalten. Dabei ergibt sich die Schwierigkeit, welche Kategorien oder Aspekte für eine solche Spezifizierung zu wählen sind, und wie eine solche Auswahl aus dem jeweiligen Gegenstandsbereich zu begründen ist. Ein solches deduktives Verfahren wird ein Relationeninventar hervorbringen, welches sich voraussichtlich durch einen höheren Grad an Allgemeinheit auszeichnet, möglicherweise aber nicht die notwendige oder gewünschte Spezifität aufweist. Auf diese Weise geschaffene Inventare eignen sich vielleicht gut für umfangreiche und wenig spezialisierte Gegenstandsbereiche oder domänenübergreifende Ansätze, verschenken aber möglicherweise wertvolles Differenzierungspotential und bergen die Gefahr in sich, mit einigen wenigen, zu allgemeinen Relationstypen genau die Art von unscharfer Erschließung zu provozieren oder fortzuführen, die sie beheben sollen. Der damit verbundene bereits erwähnte Verlust an gestalterischen Möglichkeiten, die Schwierigkeiten einer philosophischen Begründung, sowie die mangelnde Nachvollziehbarkeit des Definitionsprozesses lassen dieses Verfahren in seiner Reinform wenig geeignet erscheinen, wenngleich es einige interessante Denkansätze und Problemstellungen verdeutlicht, die auch für andere Verfahrensweisen von Interesse sein können. Inwieweit ein solcher deduktiver Ansatz in Teilen und in Kombination mit anderen Herangehensweisen durchaus nützlich sein kann, soll zu einem späteren Zeitpunkt wieder aufgegriffen werden.

Diesem Vorgehen entgegengesetzt ist eine induktive Herangehensweise, bei der zuerst hochgradig spezifische Relationen aus den speziellen Eigenschaften und Zusammenhängen der bereits vorgegebenen Termmenge herausgearbeitet und anschließend in einem Reduktions- und Auswahlprozess gleich einer Destillation allgemeinere Relationstypen gewonnen werden. Hierbei wird versucht, aus einer großen Menge sehr spezifischer Relationen, durch fortschreitende Bündelung, Abstraktion und Verallgemeinerung ein begrenztes Inventar, welches sich durch eine hinreichende Allgemeinheit bei gleichzeitig größtmöglicher Spezifität der Relationen auszeichnet. Dieser Reduktions- und Abstraktionsprozess soll im folgenden Unterkapitel dargestellt werden.

6.6.1 Reduktions- und Abstraktionsprozesse

Die Reduktion einer umfangreichen und hochgradig heterogenen Menge an Relationstypen zu einem begrenzten und strukturierten Relationeninventar kann als ein Verfahren der schrittweisen Zusammenfassung, und Abstrakti-

on von speziellen hin zu allgemeineren Relationstypen gesehen werden. Dieser Prozess der Bündelung zahlreicher hochgradig spezifischer Relationen muss intellektuell geleistet werden und lässt sich nur sehr begrenzt durch den Einsatz von Logik im Rahmen rationalisierter Prozesse ersetzen. Vor allem die Analyse und der Vergleich des semantischen Gehalts einzelner Relationstypen muss intellektuell geleistet werden.

Die Auswahl und Bündelung ähnlicher Relationen zu einigen wenigen hinreichend allgemeinen aber dennoch genau definierten Relationen sowie die anschließende Validierung dieses Prozesses ist für die Aussagekraft und die Funktionalität der mithilfe dieser Relationen gestalteten Strukturen der Dokumentationssprachen entscheidend.

Hier soll versucht werden, verschiedene Methoden und Prinzipien zur Herausbildung eines solchen Relationeninventars aus der Menge sehr spezifischer, teilweise nur zwischen zwei Begriffen existierenden Relationen, darzustellen und deren Anwendbarkeit im Rahmen der speziellen Aufgabenstellung zu diskutieren.

Im Gegensatz zu anderen Ansätzen, welche das Relationeninventar durch eine gegenstandsunabhängige, dafür aber sprachspezifische linguistische Analyse der innerhalb einer Sprache vorkommenden und unterscheidbaren semantischen Relationen bestimmen, wird in dem vorgestellten Verfahren das Relationeninventar aus einem, durch bereits vorgegebene, teils komplexer Begriffe einer bestehenden Dokumentationssprache definierten Gegenstandsbereich heraus entwickelt werden.

Ebenso wie die Bildung eines (Allgemein-) Begriffs immer das Ergebnis eines von vielen Einzelfällen und Beispielen ausgehenden Abstraktionsprozesses ist, so wird auch bei der Suche nach allgemeinen Relationen versucht, das was möglichst vielen der sehr spezifischen Relationen gemein ist, herauszuarbeiten und von den Einzelfällen zu abstrahieren, um somit eine allgemeingültige, aber dennoch für eine aussagekräftige Modellierung jeweils hinreichend spezifische Relation zu erhalten.

Die Schaffung eines umfangreichen und detaillierten Inventars dient in erster Linie nicht einer möglichst feingliedrigen, logischen Unterteilung eines Gegenstandsbereichs, sondern dazu, Zusammenhänge und Bezüge herzustellen und aufzuzeigen, die für die Orientierung und Suche in dem jeweiligen Gegenstandsbereich relevant sein können. Die so erschaffenen Strukturen stellen keine objektive Wiedergabe des Gegenstandsbereichs dar, sondern sind eine anwendungsorientierte Modellierung, welche durch ihre Funktion als Dokumentationssprache verobjektiviert wird.

Diese Überlegungen zeigen, dass nicht nur die Ausweisung der Relationen zwischen den Begriffen, sondern auch die Definition des Relationeninventars intellektuell geleistet werden muss. Sowohl die Findung der Einzelrelationen als auch der Reduktionsprozess zur Schaffung eines begrenzten Inventars lassen sich nicht auf einfache maschinell ausführbare Anweisungen reduzieren. Die Zusammenfassung und Bündelung der Einzelrelationen in ihrer semantischen Reichhaltigkeit und Verschiedenheit kann nicht aus-

schließlich aus deren logischen oder strukturellen Eigenschaften abgeleitet werden.

Trotzdem können die logischen Eigenschaften sowie die über die jeweiligen „domains“ und „ranges“ möglichen „Subjekte“ und „Objekte“ der Relationen für die Validierung der Relationen und des Reduktionsprozesses von Nutzen sein. So kann anhand der Klassen, welcher die Subjekte und Objekte angehören, überprüft werden, ob die semantische Ähnlichkeit mehrerer Einzelrelationen die zusammengefasst werden sollen, auch bezüglich ihrer Funktionsweise und ihrer Valenz übereinstimmen. Eine solche Übereinstimmung ist notwendig, da der Informationsverlust beim Austausch der spezifischen Einzelrelationen durch Inventarrelationen so gering wie möglich gehalten werden soll, und die logischen Eigenschaften erhalten bleiben müssen. Daher sollte versucht werden, jeweils nur transitive, symmetrische oder andere, aber genau definierte Einzelrelationen zusammenzufassen, um Inkonsistenzen zu vermeiden.

6.6.2 Typisierung der Relationsarten

Entsprechend dieser Überlegungen zu den Möglichkeiten einer induktiven Herangehensweise bei der Inventarbildung wurde folgendes Vorgehen ausgewählt: Sammlung aller zwischen jeweils zwei Begriffen bestehenden Relationen welche „allgemeingültig“ sind, *a priori* bestehen und einer dem Gegenstandsbereich angemessenen „Verobjektivierung“ entsprechen, und deren natürlichsprachlicher Repräsentationen in Referenzwerken oder anderen Quellen. Das bereits bestehende Relationengerüst aus der SWD wurde dabei nicht *direkt* übernommen, alle ausgezeichneten Relationen wurden einzeln auf ihre Richtigkeit geprüft und nachträglich differenziert.

Für die Sammlung und Zusammenstellung aller über die bestehende Relationierung hinausgehenden, zwischen den Begriffen der MACS-Liste bestehenden Beziehungen wurden alle Begriffe in den entsprechenden Referenzwerken⁷¹ nachgeschlagen und explizite Verweise auf andere Lemmata, aber auch in den Lexikonartikeln in Form von Beschreibungen enthaltenen Informationen auf mögliche Beziehungen zu weiteren Begriffen hin untersucht. Die darin dargestellten Beziehungen und Zusammenhänge wurden für jedes Schlagwort auf der MACS-Liste in einer ersten Relationenliste festgehalten. Diese Liste bildet die Grundlage für den bereits beschriebenen Prozess der „Inventarisierung“ durch Analyse, Abstraktion und Bündelung der hochspezifischen „rohen“ Begriffsrelationen hin zu präzise definierten „Inventarrelationen.“

⁷¹ Es wurden die in der „Liste der fachlichen Nachschlagewerke zu den Normdateien (GKD, PND, SWD) Ausgabe April 2005. Frankfurt a. M.: Die Deutsche Bibliothek 2005“ als vorrangig zu konsultierende Referenzwerke genutzt:

Manfred Brauneck; Gerhard Schneilin. Theater Lexikon. Begriffe und Epochen/Bühnen und Ensembles. Reinbek bei Hamburg: Rowohlt, 1986

Lexikon Theater International. Jochanan Ch. Trilse-Finkelstein, Klaus Hammer. Berlin: Henschel, 1995

Theater-Lexikon. Rischbieter, Henning (Hrsg.). Zürich [u. a.] Orell Füssli, 1983

Zusätzliche Orientierung kann die Überprüfung des Vokabulars auf bestimmte mehrfach auftauchende Muster sein. So kann das Vorhandensein zahlreicher Berufsbezeichnungen sowie der entsprechenden Tätigkeiten und den daraus resultierenden Produkten oder Objekten ein Hinweis auf die Notwendigkeit von Relationen sein, welche eine Modellierung dieser Zusammenhänge ermöglichen. In diesem Fall wurden diese als Agensrelationen „Agent-Tätigkeit“ und „Tätigkeit-Resultat“ definiert.

Bei diesem Prozess werden diejenigen Relationen, welche bedeutungsgleich oder in ihrem semantischen Gehalt sowie ihrer Funktion sehr ähnlich sind, zusammengefasst und in einer geringfügig verallgemeinerten neuen Relation vereint. Um bei diesem Reduktionsprozess den Verlust an Information, Spezifität der Relationen so gering wie möglich zu halten, müssen bereits während des Prozesses einige Voraussetzungen erfüllt sein. Dabei ist das größte Problem die Findung und Definition von Kriterien, welche für eine solche Zusammenfassung gegeben sein müssen. Die Forderung, die Relationen, welche zusammengefasst werden, müssen jeweils in Ihrer Domain und Range übereinstimmen, erweist sich als problematisch, da das Vokabular nicht nach logischen Prinzipien oder festen Regeln gebildet wurde, sondern dem natürlichsprachlichen fachlichen Diskurs entstammt und nach den Prinzipien der Gebräuchlichkeit ausgewählt wurde. Zwar lässt sich auch ein solchermaßen gebildetes Vokabular noch nachträglich einer groben Kategorisierung unterziehen und in verschiedene Klassen ordnen, diese Klassen oder Kategorien dann aber für die Definition von Relationen oder deren Validierung zu verwenden ist problematisch, da es sich bei den Klassen nicht um logisch valide Konstrukte, sondern um Ordnungshilfen handelt. Dass Begriffe abhängig von der sprachlichen Gestalt ihrer Bezeichnung in unterschiedlichen Klassen enden können muss auf der begrifflichen Ebene vermieden bzw. überwunden werden.

Die Unterteilung dieses Vorgehens in die zwei getrennten Schritte der Sammlung und der Bündelung mag künstlich erscheinen. Tatsächlich werden auch in vielen Fällen bereits während des Prozess des Sammelns und der Analyse bestimmte Abstraktions- und Bündelungsprozesse ablaufen und es besteht die Gefahr, dass man bereits zu diesem Zeitpunkt eine Reduktion und somit eine „Kanonisierung“ eines begrenzten Inventars möglicherweise unbewusst betreibt. Da im Rahmen dieser Arbeit aber die Überlegungen und Vorgehensweisen, die zur Bildung und präzisen Beschreibung eines solchen Inventars notwendig sind, ausführlich dargestellt werden sollen, wurde dieser Prozess bewusst in zwei Stufen unterteilt. Des Weiteren besteht bei einer unbewussten oder auch intuitiven Zusammenfassung aller relevanter Relationen die Gefahr, dass bestimmte logische Eigenschaften nicht korrekt erkannt oder nicht ausreichend dokumentiert werden, und so im Prozess der Inventarbildung nicht ausgewiesen werden, was wiederum zu Inkonsistenzen führen kann.

Daher sollte an dieser Stelle bewusst der Gebrauch allzu allgemeiner Relationen nach Möglichkeit noch vermieden werden. Ein mögliches Vorge-

hen, um dies zu verhindern, ist die direkte Umsetzung von den in Referenzwerken in Form von natürlichsprachlichen Sätzen gefundenen Aussagen in formalisierte Aussagen wie „Regisseur bestimmt Rollenbesetzung“ und die anschließende Dekonstruktion des Satzes. Das Herauslösen der Verbalphrase, die die Beziehung zwischen den beiden Begriffen aus der Liste bezeichnet, ermöglicht eine erste unmittelbare Findung spezifischer Begriffsrelationen. Die so entstandenen Sätze zeichnen sich durch eine einfache, dreigliedrige Struktur und einen mittleren, bereits deutlich erkennbaren Formalisierungsgrad aus. Sie bestehen aus Subjekt, Prädikat und Objekt bzw. Subjektkomplement.

In diesen Sätzen sind Subjekt und Objekt Elemente aus dem kontrollierten Vokabular des Begriffsinventars der MACS Liste, das jeweilige Prädikat stellt die Art der Beziehung möglichst detailliert und in natürlichsprachiger Form dar. Um das oben beschriebene Phänomen einer bereits zu diesem Zeitpunkt unbewusst vorgenommenen Abstraktion zu unterbinden, ist es empfehlenswert, sich bei der Formulierung des Prädikats möglichst eng an den Texten in den Referenzwerken zu halten. Beziehungen, welche in den Referenzwerken nur durch nicht näher spezifizierte Verweise dargestellt werden, können entweder direkt natürlichsprachlich ergänzt werden, oder die Relationen bleiben undefiniert und werden zu einem späteren Zeitpunkt durch Relationen aus dem dann gebildeten Inventar ergänzt.

In einem zweiten Schritt werden diese Ausgangssätze verglichen, um Ähnlichkeiten bei den verwendeten Verbalphrasen und den von ihnen dargestellten Relationen zu finden und zu dokumentieren. Dabei wird versucht, wenn möglich die ersten natürlichsprachlichen Bezeichnungen durch ähnliche sprachliche Konstruktionen zu ersetzen, um die Suche nach semantisch äquivalenten Relationen sowie den Prozess der Bündelung, der in einem folgenden dritten Schritt zur Erstellung des reduzierten Inventars benutzt wird, zu erleichtern. Hierbei sollte allerdings darauf geachtet werden, dass etwaige, unbewusst ablaufende Abstraktionsprozesse sowie Vereinfachungen vermieden werden oder zumindest soweit dies möglich ist, transparent dargestellt werden, um sie später dokumentieren und zum Gegenstand der Analyse und Validierung machen zu können. Bei dem ersten Schritt der sprachlichen Vereinheitlichung sollte darauf geachtet werden, dass die Verwendung bestimmter verallgemeinerter Konstruktionen oder Begriffe nicht die Bedeutung, also den semantischen Gehalt und die logischen Eigenschaften des Prädikats verändern, da diese den semantischen Kern und damit das Ausgangsmaterial für den darauf folgenden Reduktionsprozess bilden.

Diese einfachen Aussagesätze deren Subjekt und Objekt bzw. Subjektkomplement aus jeweils einem Term des kontrollierten Vokabulars und deren Prädikate aus einfachen Verben oder Verbalphrasen bestehen, ähneln in ihrer Form den in RDF-Konstrukten verwendeten Tripeln, aus Ressourcen (*resources* oder *instances*) und Eigenschaften (*properties*) und nehmen damit schon bestimmte strukturelle Eigenschaften der Modellierungssprache auf, was eine spätere Umsetzung in der Softwareumgebung erleichtert, ob-

schon bedacht werden muss, dass eine solche Modellierung anhand einer spezifischen Ontologie-Sprache oder eines Modellierungsframeworks immer nur mit dem fertigen „Relationeninventar“ gestaltet werden wird.

6.6.2.1 Logische Charakterisierung

Möglichkeiten und Grenzen der Darstellung logischer Eigenschaften von Relationen hängen eng mit deren semantischem Gehalt, aber auch mit den Begriffstypen welche sie verbinden sollen, zusammen.

Neben der zwar teilweise natürlichsprachlichen aber dennoch präzisen und formalisierten Definition des semantischen Gehalts einer Relation ist auch der Ausweis ihrer logischen Eigenschaften für deren Verwendung in Ontologien von Bedeutung. Vor allem die logischen Eigenschaften der Transitivität und Symmetrie sowie der Ausweis der jeweils inversen Relationen ist für die präzise Definition semantischer Relationen in Dokumentationssprachen, welche das Information Retrieval unterstützen sollen wichtig. Während die Transitivität hierarchischer Relationen sowie deren formale Auszeichnung beispielsweise eine wichtige Voraussetzung für die bereits beschriebene Expansion von Suchanfragen ist, können andere Angaben wie beispielsweise die Definition bestimmter Relationen als funktional oder invers funktional dem Benutzer zusätzliche Informationen über den modellierten Gegenstandsbereich geben und seine Orientierung verbessern. Voraussetzung dafür ist jedoch, dass die logischen Eigenschaften transparent und nachvollziehbar dargestellt werden.

Aufgrund der hohen Spezifität der Relationen und der Tatsache, dass vor allem durch nichthierarchische Relationen unterschiedliche Begriffstypen verknüpft werden, sind nur für einen kleinen Teil dieser Relationen besondere logische Eigenschaften ausweisbar. Dagegen sind die für die beschriebenen Retrievalfunktionalitäten wichtige logische Eigenschaft der Transitivität sowie die damit verbundene Vererbbarkeit bei den hierarchischen Relationen vorhanden und entsprechend ausgewiesen.

Wegen dieser Einschränkungen lassen sich einige logische Eigenschaften oder die Definition von Verwendungseinschränkungen lediglich als Hilfestellungen und Kontrollmöglichkeiten während der Modellierung nutzen. So ist beispielsweise die Definition der Äquivalenzrelation „Benutze Synonym“ (EQU/USE) als invers funktional zur terminologischen Kontrolle sinnvoll, damit Synonyme präzise *einer* einzigen Vorzugsbenennung zugeordnet und Mehrfachzuweisungen oder uneindeutige Zuordnungen von zwei oder mehr Vorzugsbenennungen vermieden werden können.

Auch bei der Definition der jeweiligen *Domains* und *Ranges* der einzelnen Relationen ist eine Unterteilung der Begriffe in hierarchisch zueinander geordnete Klassen möglicherweise von Vorteil, dabei sollte aber darauf geachtet werden, dass die dadurch entstehenden Einschränkungen bei der Modellierung von Begriffszusammenhängen keine sinnvollen Modellierungen verhindern.

Der Gebrauch von Klassen zur Strukturierung von Begriffsmengen sowie zur semantischen Modellierung von Ontologien bietet auch im Fall terminologischer Ontologien einige zusätzliche Ordnungs- und Gestaltungsmöglichkeiten.

Bei der Modellierung in RDF/OWL ist allerdings zu beachten, dass Relationen nur zwischen Instanzen nicht aber zwischen Klassen ausgewiesen werden können. Daher ist die Strukturierung des Vokabulars durch eine monohierarchische Klassenstruktur durchaus möglich, die jeweiligen Begriffe, die zur Definition einer Klasse herangezogen werden und deren Beziehungen zueinander die hierarchische Klassenstruktur bestimmen, müssen allerdings auch als Instanzen definiert werden um ihre vollständige Relationierung zu ermöglichen. Das Resultat wäre eine Ontologie bei welcher die Klassenstruktur die systematische Ansicht, die zwischen den Instanzen ausgewiesenen Relationen die Darstellung der semantischen Zusammenhänge bietet.

Für ein leistungsstarkes Retrieval ist eine Differenzierung, sowie eine Austypisierung der Relationen eine wichtige Voraussetzung. Die Differenzierung des Relationeninventars sowie die daraus erwachsende Heterogenität können aber auch Probleme mit sich bringen. Bei klassischen Thesauri, bei denen der gesamte Begriffsraum durch einige wenige Relationstypen strukturiert ist, ist vor allem durch die hierarchischen Relationen ein durchgängiger Zusammenhang zumindest für den Großteil des Termininventars gewährleistet, auch wenn durch verschiedene Ansichten oder Darstellungsoptionen nur einzelne Relationstypen für die Darstellung der Strukturen ausgewählt werden. Dies gilt im Besonderen für die hierarchischen Relationen, welche gleichsam ein stützendes Gerüst für den gesamten Thesaurus bereitstellen. Werden diese hierarchischen Relationen wie empfohlen austypisiert und zu einer differenzierten Relationierung eines Vokabulars herangezogen, kann dies zu Problemen bezüglich der Übersichtlichkeit des Vokabulars und der Orientierung im Rechercheprozess führen: Werden, ausgehend von einem ausgewählten Begriff, nur einzelne sehr spezifische Subtypen der hierarchischen Relationen ausgewählt, kann dies zur Darstellung sehr kleiner inselartiger Ausschnitte aus dem Vokabular führen. Dies ist für die punktuelle Betrachtung solcher Ausschnitte und bei der Formulierung entsprechend spezifischer Suchanfragen sicherlich von Vorteil, birgt aber die Gefahr, dass der Zusammenhang mit dem restlichen Vokabular und anderen möglicherweise relevanten Bereichen zu früh aus dem Blick gerät.

Um die Auswahl der angezeigten Relationen und somit die Spezifizierung der dargestellten Bereiche schrittweise und flexibel gestalten zu können, sollten ähnliche Relationen, welche bestimmte logische Eigenschaften teilen oder zu bestimmten Gruppen gehören auch gemeinsam ausgewählt werden können, um gerade für eine anfänglich möglicherweise gewollt unspezifisch gestaltete Suche, möglichst reichhaltige semantische Nachbarschaften betrachten zu können. So kann eine gemeinsame Anzeige aller vorhandenen hierarchischen Relationen durchaus sinnvoll sein, um dann in

einem zweiten Schritt, nach Sichtung der damit verbundenen Terme eine Eingrenzung über eine Spezifizierung dieser Relation nach bestimmten Aspekten vorzunehmen. Diese Funktionalität kann durch einen hierarchisch strukturierten Aufbau des Relationeninventars und die Einrichtung entsprechender Auswahlmöglichkeiten erreicht werden.

Daher ist ein weiterer, optionaler aber durchaus sinnvoller Schritt bei der Gestaltung des Relationeninventars, dessen hierarchische Strukturierung.

6.6.2.2 *Aufbau einer hierarchischen Struktur / Taxonomie*

Der Begriff der Taxonomie erfolgt hier in Anlehnung an das von Dee⁷² im Rahmen einer ALA Studie⁷³ entwickelte und als „Taxonomy of Relationships“ bezeichnete Relationeninventar. Der Begriff der Taxonomie soll hier nicht in seiner streng philosophischen Bedeutung verstanden werden sondern zur Bezeichnung einer einfachen hierarchischen Struktur zu Zwecken der Ordnung und übersichtlichen Darstellung einer begrenzten Menge zueinander geordneter Elemente -in diesem Fall Relationstypen - verstanden werden.

Eine solche Taxonomie bietet einen gestuften, hierarchischen Aufbau, in dem bestimmte Eigenschaften von Relationen vererbt werden können. In der Modellierung könnte dies auch durch eine zwei- oder mehrgliedrige Definition und Bezeichnung der entsprechenden Relationen verdeutlicht werden.

Im Fall der hierarchischen Relationen die bisher nur durch die Begriffsbeziehung OB/UB dargestellt wurde, könnte eine solche stufenweise Differenzierung wie folgt aussehen:

OB
OB / generisch
OB / generisch / Aspekt: Funktion

Dies entspricht einer hierarchisch geordneten Relationentaxonomie, bei welcher die jeweiligen semantischen und logischen Eigenschaften von dem jeweils übergeordneten Relationstypen oder Eigenschaften (*superproperties*) an die ihnen untergeordneten Untereigenschaften (*subproperties*) ver-

⁷² Dee, Michel, Taxonomy of subject relationships, Appendix B zu: American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSSSubject Analysis Committee. (1997)

⁷³ American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSSSubject Analysis Committee. (1997)

erben, und dies auch entsprechend formal streng dargestellt und maschinenlesbar hinterlegt ist.

Ebenso können die beiden anderen Grundtypen von Relationen, die nicht-hierarchischen und die Äquivalenzrelationen, durch immer weitergehende Spezifizierungen schrittweise differenzieren und in einer abgestuften Hierarchie strukturiert zueinander angeordnet und übersichtlich dargestellt und verwendet werden.

Bei diesem Vorgehen können sich die an anderer Stelle in dieser Arbeit beschriebene induktive und deduktive Entwicklungs- und Bildungsmethode für Relationeninventare ergänzen: Einerseits werden aus der Menge der Einzelrelationen ähnliche oder semantisch gleiche Relationen zusammengefasst, das heißt von den speziellen, vielleicht sogar einmaligen Relationen hin zu allgemeineren Relationstypen gebündelt. Andererseits werden diese, aus den zusammengefassten Einzelrelationen mit ihrer maximalen Spezifität herausgearbeiteten und von den einzelnen Prädikaten der primitiven Aussagen der einzelnen Sätze abstrahierten Relationen als spezifische „Unterrelationen“ der drei großen Grundtypen verstanden. So werden sie entsprechend ihrer semantischen Spezifität auf verschiedenen Stufen in die so entstehende Relationentaxonomie eingeordnet.

Die differenzierte Gestaltung der Relationen sowie die hierarchisch-taxonomische Ordnung des gesamten Relationeninventars durch eine mehrstufige und möglicherweise modular aufgebaute und daher bei Bedarf erweiterbare Binnengliederung stellt besondere Anforderungen an die Navigations- und Visualisierungskomponenten. Sie bietet aber bei einer angemessenen und sinnvollen Umsetzung in entsprechende Darstellungs- und Interaktionsformen einige entscheidende Vorteile:

Die hierarchische Gliederung des Inventars bietet dem Benutzer die Möglichkeit, den Differenzierungsgrad der zur Navigation und Frageformulierung benutzten Relationen seinen Bedürfnissen und Fragestellungen entsprechend zu wählen. Für einen ersten Einstieg und einen groben Überblick können die bekannten und möglicherweise vertrauten Standardrelationen OB/UB, VB, BF/BS benutzt werden, die dann nach einer ersten Orientierung zwecks einer differenzierteren Navigation verfeinert werden können. So kann zum Beispiel eine Begriffsumgebung, welche über die Standardrelation NT zu viele oder zu heterogene Unterbegriffe enthält, durch die gezielte Auswahl einzelner, spezifischerer Subtypen nach bestimmten Gesichtspunkten gegliedert und für die Erkundung übersichtlich gestaltet werden.

Der modulare Aufbau bietet die Möglichkeit, die allgemeineren Relationen auf den oberen Hierarchieebenen der Relationentaxonomie für das gesamte Vokabular einer Dokumentationssprache zu definieren und zu verwenden und bei Bedarf spezifischere Relationen für die Relationierung eines bestimmten Gegenstandsbereichs an die unterste Hierarchieebene anzukoppeln um so eine Erweiterung anhand einer noch weitergehenden Spezifi-

zierung des Inventars für bestimmte Gegenstandsbereiche oder Spezialfälle zu erreichen.

Solche, für eine Erweiterung erstellten Relationen können als eigenständig definierte Relationen ausgewiesen werden oder sie können lediglich für die Relationen, an welche sie angekoppelt werden den Aspekt nachdem eine Differenzierung der Relation und damit die Verknüpfung der Begriffe vorgenommen wird, spezifizieren.

Als Beispiel hierfür kann die weitere Unterteilung und Präzisierung der generischen Relation dienen, welche durch die Nennung verschiedener Aspekte und deren Nutzung zur Definition der spezifischeren Relationen, wie sie in der terminologischen Ontologie für den Gegenstandsbereich Theater definiert sind, erreicht werden können.

Die allgemeine hierarchische Relation wird auf der zweiten Ebene in die partitive und die generische Relation unterteilt. Eine weitere Unterteilung der generischen Relation findet auf der dritten Ebene nach verschiedenen Aspekten statt, welche sich immer noch durch eine Allgemeinheit auszeichnen, welche die Verwendung dieser differenzierten Relationen für die Relationierung des gesamten Vokabulars ermöglicht. Wird eine weitere Differenzierung gewünscht so ist dies auf einer zusätzlichen vierten Ebene möglich. Da eine solche sehr weitgehende, durchaus sinnvolle, Differenzierung in hohem Maße vom Gegenstandsbereich abhängig ist, erscheint die Definition verschiedener domänenspezifischer Module, welche die spezifischen Relationen enthalten sinnvoll. So ist im Beispielfall eine weitere Differenzierung der nach dem Aspekt „Beteiligte“ unterschiedenen generischen Relation in „Publikum“ und „Aufführende“ auf der angekoppelten, vierten und in diesem Fall letzten Ebene möglich.

1. Ebene	2. Ebene	3. Ebene	4. Ebene (Theater) (angekoppelt)
OB			
	OBG generisch		
		OBG / Beteiligte	
			OBG / Beteiligte / Aufführende
			OBG / Beteiligte / Publikum

		OBG / Funktion	
		OBG / Form	
			OBG / Form / Genre
			OBG / Form / Tradition
		OBG / Lokaler Bezug	
		OBG / Methode	
		OBG / Instituti- on	

So wäre für einen anderen Gegenstandsbereich die Differenzierung der generischen Relation „*OBG / Beteiligte*“ nach anderen Aspekten sinnvoll.

Ob die angekoppelten Module jeweils nur aus einer einzigen, letzten Hierarchieebene bestehen oder auch hier eine mehrstufige Binnengliederung sinnvoll ist, hängt ebenfalls von dem zu modellierenden Gegenstandsbereich ab. Ebenfalls domänenspezifisch ist die Frage welche Relationen durch weitere Relationen auf einer oder mehreren zusätzlichen Hierarchieebene differenziert werden sollen. Grundsätzlich besteht eine modulare Erweiterbarkeit aber bei allen Relationen der jeweils letzten, spezifischsten Hierarchieebene.

Die Entwicklung eines solchen Inventars unterstützt auch die Erfüllung eines weiteren Desiderats in Bezug auf die Relationierung der SWD. Ein umfangreicheres und differenziertes Relationeninventar kann die notwendige Verdichtung der bestehenden Relationierung, das Schließen von Lücken sowie die Einbindung bislang gänzlich unverbundener und daher für eine Navigation sowie ein Retrieval, das auf eben diesen Strukturen aufsetzen, nicht zu nutzenden Terme erleichtern.

Diese „alleinstehenden“ Terme, die zwar in der SWD erhalten sind und für die Erschließung genutzt werden können, da sie dem Indexierer aufgrund seiner profunden Kenntnis des Gegenstandsbereichs bekannt sind, vom Benutzer der sich einen Gegenstandsbereich erst explorativ erschließen muss, aber nur durch Zufall gefunden und für eine Suche genutzt werden können, sind für das Retrieval „verloren“. Daher sind unrelationierte Terme in Thesauri oder anderen Dokumentationssprachen zu vermeiden. Die Verwendung eines differenzierten Relationeninventars kann dabei helfen, eine passende und in ihrer semantischen Aussage sinnvolle Relation zur Einbindung eines solchen isolierten Terms zu modellieren, ohne dabei auf eine unpräzise oder sogar inkorrekte Notlösung wie die Ausweisung einer unspezifischen „assoziativen Relation“ zurückgreifen zu müssen. Durch ein differenziertes, idealerweise dem Gegenstandsbereich angepasstes Inventar können

selbst sehr spezifische Begriffe präzise und aussagekräftig in Begriffsstrukturen integriert werden.

Folgende Relationstypen wurden in das Inventar aufgenommen:

- hierarchische Relationen
- nichthierarchische Relationen
- Äquivalenzrelationen

Diese wurden wie folgt differenziert und in spezifischere Relationen aufgeteilt:

Die zusätzliche und künstliche Aufteilung der hierarchischen Relationen in zwei separate, zueinander inverse, gerichtete Relationen „*Oberbegriff*“ und „*Unterbegriff*“ erklärt sich aus dem Versuch, auf der obersten Ebene der Relationentaxonomie die bereits bekannten und für die Relationierung der SWD verwendeten Relationen sowie deren Bezeichnungen zu übernehmen, um so dem Benutzer zumindest für den Einstieg auf der allgemeinsten Ebene einen Anknüpfungspunkt an bereits Bekanntes zu bieten. Bei den nicht-hierarchischen Relationen wurde hingegen bewusst darauf verzichtet, die ungenaue und in ihrer mangelnden Spezifität möglicherweise irreführende Bezeichnung „Verwandter Begriff“ zu verwenden, wie dies bereits dargelegt wurde. So zeigt sich für den Ausschnitt aus der Taxonomie für den Bereich der hierarchischen Relationen folgendes Bild:

Hierarchische Relation
Oberbegriff
Oberbegriff generisch
Oberbegriff generisch nach Institution
Oberbegriff generisch nach lokalem Bezug
Oberbegriff generisch nach Methode
Oberbegriff generisch nach Professionalität
Oberbegriff generisch nach Thema
Oberbegriff generisch nach Beteiligten
Oberbegriff generisch nach Form
Oberbegriff generisch nach Funktion
Oberbegriff partitiv
Oberbegriff partitiv / integral
Oberbegriff partitiv / kontributiv
Unterbegriff
Unterbegriff generisch
Unterbegriff generisch nach Beteiligten
Unterbegriff generisch nach Form
Unterbegriff generisch nach Funktion
Unterbegriff generisch nach Institution
Unterbegriff generisch nach lokalem Bezug

Unterbegriff generisch nach Methode Unterbegriff generisch nach Professionalität Unterbegriff generisch nach Thema Unterbegriff partitiv Unterbegriff partitiv / kontributiv Unterbegriff partitiv / integral
--

Ebenso die nicht-hierarchischen Relationen:

Nicht-hierarchische Relation Affinität bevorzugte / typische Agenten bevorzugte / typische Form bevorzugtes / typisches Thema Agent Agent <--> Gegenstand Agent <--> Instrument Agent <--> Tätigkeit Agent <--> Wirkungsstätte Beitrag Beteiligte Entwicklung Differenzierung Ersetzung Hervorbringung Fokus Gestaltung Erschaffung Regelung Verantwortlichkeit Kontext Kontext entwicklungsgeschichtlich Kontext geographisch Kontext morphologisch Kontext zeitlich Tätigkeit Tätigkeit <--> Agent Tätigkeit <--> Instrument Tätigkeit <--> Objekt Tätigkeit <--> Resultat Tätigkeit <--> Ziel Thema bevorzugtes Thema implizites Thema

vorgegebenes Thema
Wirkung
Wirkung (inhibitiv)
Wirkung (kausal)
Wirkung (kontributiv)

Hier wurde bewusst auf die Verwendung der bestehenden Bezeichnung des „Verwandten Begriffs“ verzichtet, da dies nicht den unterschiedlichen semantischen Gehalten und Aussagen der Relationen entspricht.

Die Äquivalenzrelationen wurden ebenfalls um mehrere spezifischere Relationen ergänzt, welche gerade im Fall der Begriffsbeziehung „Antonym“ und „begriffliche Abgrenzung“ nur im weitesten Sinne unter diesem Begriff zu subsumieren sind. Gerade bei den beiden genannten Relationen ist zwar bei deren Verwendung im Retrievalprozess Vorsicht geboten,⁷⁴ zur Orientierung des Benutzers sowie zur Klärung unspezifischer Informationsbedarfe oder Unklarheiten über die verwendete Terminologie sind sie jedoch von Nutzen.

Äquivalenzrelation
Antonym
Begriffliche Abgrenzung (symmetrisch)
Benutze Synonym
Benutzt für
Feminine Form
Maskuline Form

6.6.2.3 Weitergehende Verallgemeinerung

Da das Relationeninventar in seiner größtmöglichen Spezifität aus dem Gegenstandsbereich heraus entwickelt worden ist, stellt eine weitergehende Verallgemeinerung einen wichtigen Schritt dar, um die Anwendbarkeit des Relationeninventars auf weitere Gegenstandsbereiche des Vokabulars oder sogar die gesamte SWD zu ermöglichen. Hierbei kann der Versuch der Relationierung eines weiteren, ebenfalls sinnvoll eingegrenzten Gegenstandsbereichs wertvolle Hinweise zur Eignung des Relationeninventars sowie auf etwaige Besonderheiten oder Defizite in Hinblick auf die Verwendung des Inventars in anderen Bereichen geben. Dadurch können in ihrer Spezifi-

⁷⁴ Zur Verwendung von Antonymen zur Sucherweiterung vgl. Wang, Yih-Chen; James Vandendorpe, Matha Evens. „Relational Thesauri in information retrieval“. In: Journal of the American Society of Information Science. 36 (1) (1985) S. 15-27.

tät sinnvoll definierte Relationstypen bestätigt werden, indem ihre erfolgreiche Verwendung in einem weiteren Sachgebiet bestätigt und entsprechend dokumentiert werden kann.

Genauso kann auch die Feststellung, dass sich einige sehr spezifische Relationstypen *nicht* für eine Anwendung in anderen Bereichen eignen, auf ein Potenzial zu einer weiteren Verallgemeinerung der Relation hinweisen, oder die Zusätzliche Schaffung weiterer Relationen nahe legen. Aufgrund der oben beschriebenen hierarchischen Struktur des Inventars sowie der durch seine Modularität gegebenen Erweiterbarkeit, können dabei Relationen, welche sich als zu spezifisch für ein allgemeines, für die Relationierung der gesamten SWD anwendbares Relationeninventar erweisen, innerhalb der Taxonomie auf die bewusst gegenstandsspezifisch gestaltete, zusätzliche und letzte hierarchische Ebene verschoben werden. So wurde beispielsweise die Relation „UBG / Aspekt: Genre“ als zu spezifisch für die Relationierungsarbeit in anderen Gegenstandsbereichen erkannt und als eine Spezifizierung der Relation „UBG / Aspekt: Form“ neu definiert und dieser auf der vierten Hierarchieebene untergeordnet, wo sie als eine gegenstandsspezifische Spezialrelation für den Bereich „Theater“ durchaus sinnvoll eingesetzt werden kann.⁷⁵

6.7 Darstellung und Anwendung der neuen Strukturen unter Protégé

Im Folgenden soll anhand einiger Darstellungen aus der Protégé-Umgebung angedeutet werden, in welcher Weise die beschriebenen Strukturen durch eine geeignete Visualisierung dargestellt werden und welche Nutzungsszenarien und Recherchestrategien durch die Neugestaltung unterstützt werden.

Die von dem Programm „Protégé“ verwendete Visualisierungskomponente „Jambalaya“ bietet eine einfache Darstellung der modellierten Strukturen sowie einige grundlegende Suchfunktionen. Die durch eine Suchanfrage gefundenen Terme werden als Instanzen nebst Ihrer unmittelbaren Nachbarn als Knoten, die sie verbindenden Relationen als gerichtete Kanten dargestellt. Die Sichtung, Auswahl und selektive Verwendung der verschiedenen Relationstypen erfolgt über ein Auswahlmenü, über das die durch verschiedenfarbige Kanten dargestellten Relationstypen einzeln oder auch frei kombiniert für die Anzeige ausgewählt werden können.

So lässt sich beispielsweise für einen Begriff das ihn unmittelbar umgebende Begriffsfeld unter der Verwendung verschiedener Relationstypen von unterschiedlichem Allgemeinheitsgrad darstellen. Werden zunächst alle oder vorwiegend allgemeine Typen ausgewählt so zeigt sich ein möglicherweise umfangreiches aber unspezifisches Umfeld. Dieses lässt sich dann schrittweise durch die Auswahl geeigneter Relationen differenzierter dar-

⁷⁵ Das gesamte Relationeninventar inklusive der speziell für den Gegenstandsbereich „Theater“ angehängten spezifischeren vierten Hierarchieebene kann in Form eines Thesaurus (Datei „relationeninventar.mth“) oder in der „Properties“-Ansicht des Projektes SWDTheater-neu (Datei „SWDTheater-neu.pprj“) betrachtet werden.

stellen und bestimmten Fragestellungen oder Informationsbedarfen entsprechend verändern.

Jeder Begriff lässt sich durch einen einfachen Befehl „Expand“ oder „Focus on“ in den Mittelpunkt rücken und zum Ausgangspunkt einer weiteren Darstellung des jeweiligen begrifflichen Umfelds – der semantischen Nachbarschaft – machen. Ein solches Vorgehen ermöglicht eine schrittweise erweiterbare Orientierung innerhalb des Vokabulars und eine intuitive und präzise gestaltbare Erkundung des gesamten Bereichs. Da sich bei der Interaktion mit diesem System zahlreiche inhaltliche Aspekte durch die Auswahl sowohl allgemeiner als auch schrittweise präzisierter, spezifischer semantischer Relationen formulieren lassen, sind sowohl eine sehr gezielte Navigation sowie auch ein eher „explorativ“ orientiertes Erkunden möglich.

Der Wechsel zwischen überblickartigen Gesamtansichten und präzise formulierten Einzelbezügen ist jederzeit über die Auswahl der Relationstypen, und die Angabe, über wie viele Knoten die entsprechende Ausweitung erfolgen soll möglich.

Die folgenden Abbildungen zeigen exemplarisch verschiedene Darstellungsmöglichkeiten mit jeweils unterschiedlich ausgewählten Relationstypen.

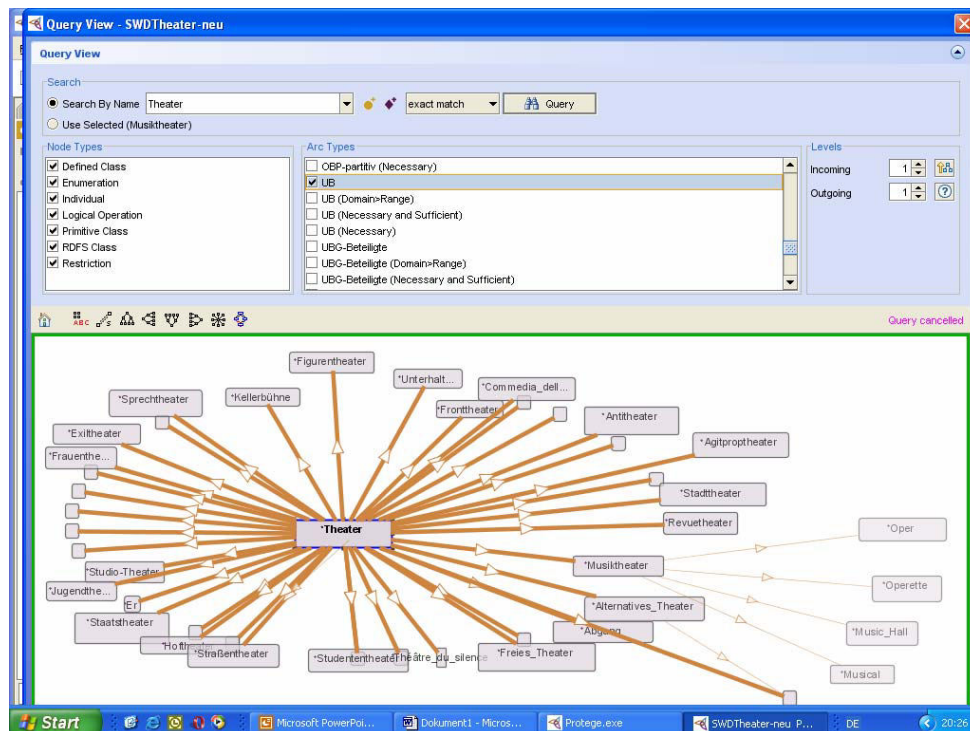


Abbildung 5: unspezifische Relationierung mittels eines Relationstyps

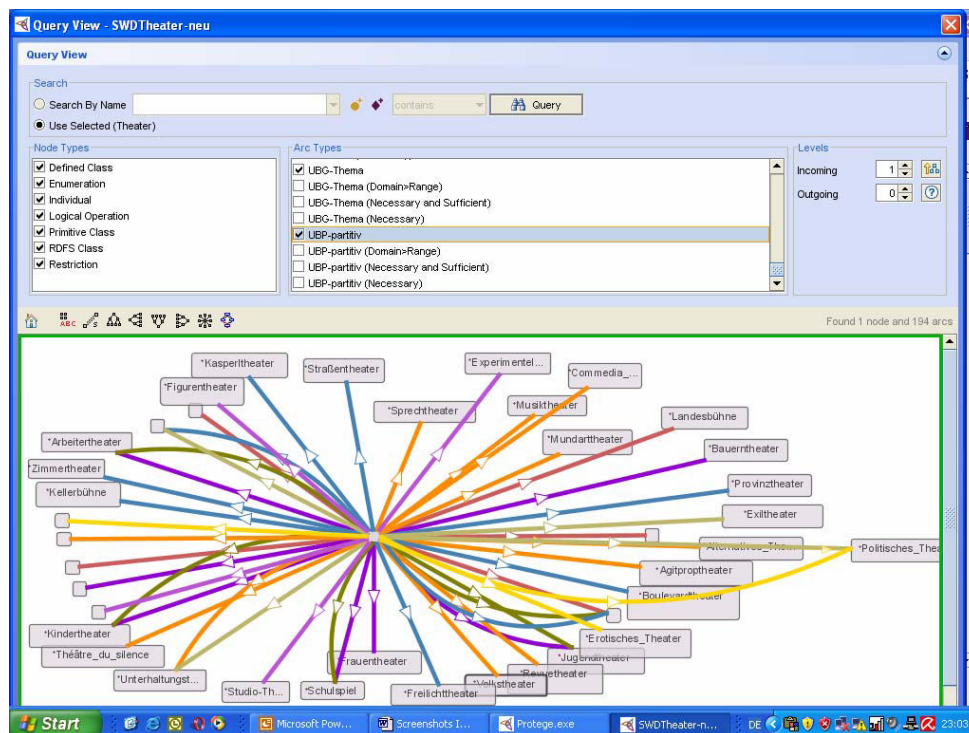


Abbildung 6: differenzierte Relationierung des Terms „Theater“ durch neue Relationstypen

Abb. 5 zeigt alle mit dem allgemeinen Einstiegsbegriff „Theater“ durch die nicht weiter spezifizierte Begriffsbeziehung „Unterbegriff“ verbundenen Begriffe, ohne diese genauer nach Art der Beziehung oder dem berücksichtigten Aspekt zu unterscheiden.

Im Gegensatz dazu zeigt Abb. 6 die gleiche Begriffsumgebung allerdings mit differenzierten Begriffsrelationen, welche durch unterschiedliche Farbgebung und genaue Bezeichnung nach dem zur Bildung des Unterbegriffs berücksichtigten Aspekt unterschieden sind und einzeln ausgewählt werden können.

Die folgenden beiden Abbildungen (Abb. 7 und Abb. 8) zeigen jeweils einzelne, gezielt ausgewählte, spezifische Relationen und verdeutlichen so den Gewinn an Übersicht und Orientierung und die Möglichkeit, bestimmte semantisch differenzierte Begriffsbeziehungen zur Erkundung weiterer Begriffe, welche möglicherweise zur Modifikation von Suchanfragen ausgewählt und eingesetzt werden können. Abb. 9 zeigt die vielfältigen, komplexen Begriffsbeziehungen welche anhand einer differenzierten Relationierung möglich sind am Beispiel des Deskriptors „*Bühnenbeleuchtung*“.

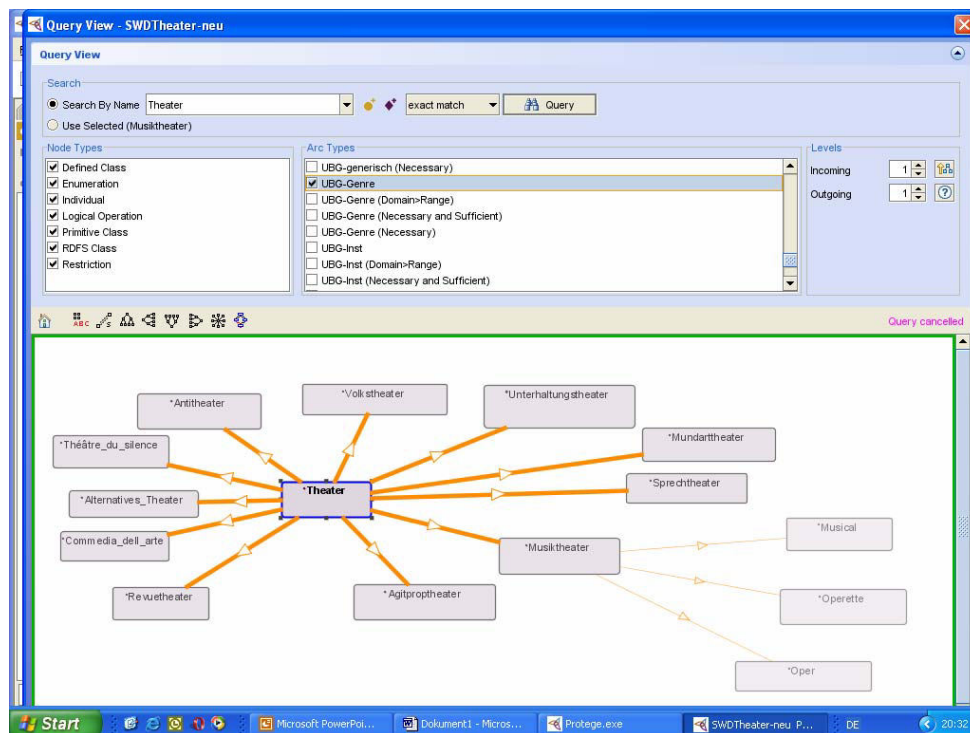


Abbildung 7: spezifische Auswahl eines Relationstyps

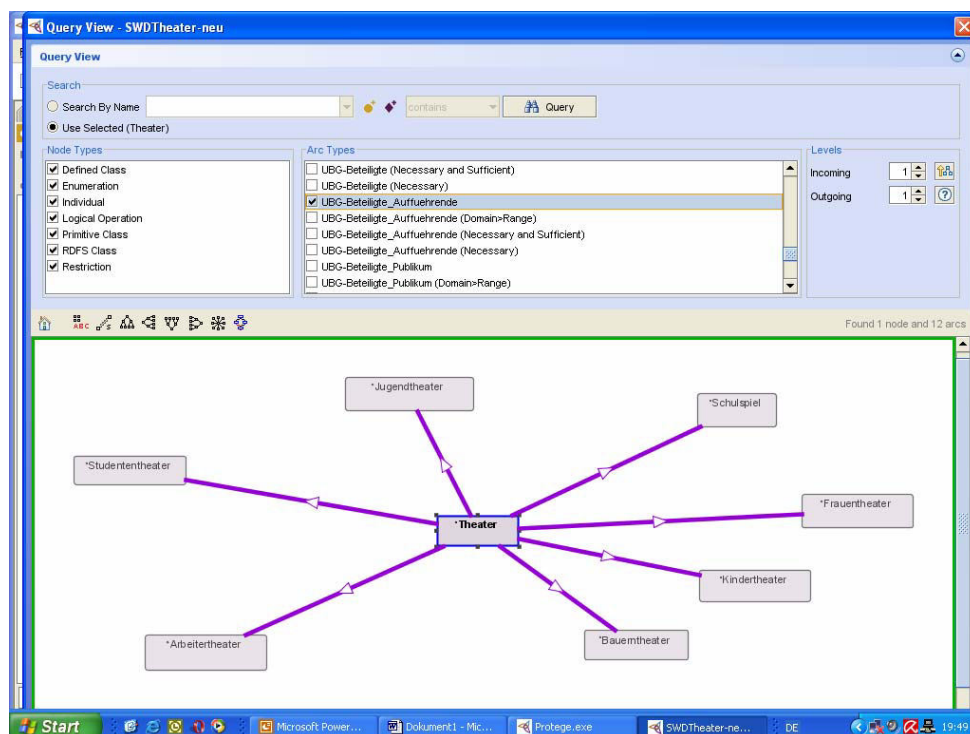


Abbildung 8: spezifische Auswahl eines Relationstyps

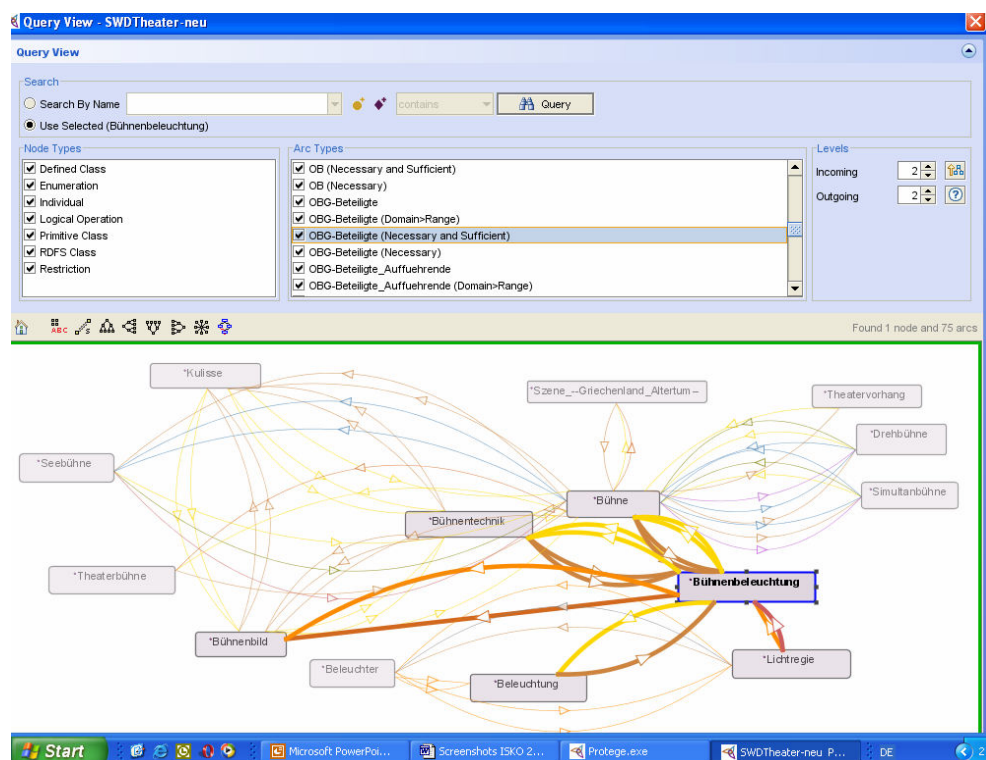


Abbildung 9: unmittelbares semantisches Umfeld des Begriffs „Bühnenbeleuchtung“

Weiter Ansichten verschiedener Suchanfragen finden sich im Anhang oder können aus der auf CD vorliegenden Modellierung dynamisch erzeugt werden.

Die Relationierung des Gegenstandsbereichs wurde in dem entsprechenden Protégé-Projekt nur selektiv für einzelne Schlüsselbegriffe sowie deren Umfeld realisiert.

Die verwendete Visualisierung sowie die Suchfunktionen haben ausschließlich Illustrationscharakter, eine verbesserte Visualisierung sowie die Gestaltung der verschiedenen Interaktionsmöglichkeiten des Benutzers mit dem System müssen für entsprechende Anwendungen anspruchsvoller entwickelt werden. Die Konzeption weiterer Komponenten und Produkte wird im sechsten Kapitel kurz thematisiert.

6.8 Regeln für die Anwendung des Relationeninventars

Um die mit einem solchen Relationeninventar durchzuführende Strukturierung des gesamten Vokabulars einheitlich und konsistent zu gestalten, werden Richtlinien zur Verwendung dieses neugeschaffenen Relationeninventars benötigt.

Ebenso wie bei der Konzeption und Erstellung des Relationeninventars bieten sich bei der Auszeichnung der jeweiligen Relationen zwischen den

einzelnen Termen verschiedene Herangehensweisen an. So kann, aufbauend auf den vorhandenen Strukturen deduktiv eine Differenzierung der bereits vorhandenen Relationen erfolgen, welche lediglich die bestehenden allgemeinen Relationen durch spezifischere Relationen ersetzt und so die bereits zwischen zwei Begriffen bestehende Begriffsbeziehung präzisiert.

Hierbei ist zu überlegen, ob eine solche Präzisierung der konkreten Relation nur entlang der Hierarchieleitern der Relationentaxonomie erfolgen können soll, oder ob auch eine freie Auswahl aus dem gesamten Inventar erlaubt sein soll. Im ersten Fall könnte eine bestehende hierarchische Relation nur durch eine spezifischere hierarchische Relation, nicht aber durch eine nicht-hierarchische Relation ersetzt werden. Dagegen würde die Möglichkeit einer freien Auswahl aus dem Relationeninventar die gestalterische Freiheit erhöhen und etwaige, durch das erweiterte und veränderte Relationeninventar notwendig gewordene Korrekturen ermöglichen.

Zusätzlich zur Präzisierung bestehender Relationen bietet die Ausweisung ergänzender neuer Relationen die Möglichkeit, Lücken in der Relationierung zu schließen und das Netz bestehender Relationen zu verdichten.

Generell besteht jedoch bei einem solchen Vorgehen die Gefahr einer zu großen Verhaftung an bestehenden Strukturen und einer unbewussten Zurückhaltung bei der Korrektur oder gänzlichen Löschung bestehender Beziehungen sowie bei der Ausweisung der durch neue Relationen ermöglichten Differenzierungen. Da bei der Relationierung immer bewusst auf eine größtmögliche Differenzierung bei der Gestaltung des Relationengeflechts geachtet werden sollte, empfiehlt sich ein alternatives Vorgehen:

Eine größtmögliche gestalterische Freiheit bei der Modellierung der neu zu schaffenden terminologischen Ontologie anhand des neuen Relationeninventars lässt sich erreichen, wenn alle Terme von Grund auf neu relationiert werden. Dabei können aus dem gesamten Inventar die jeweils angemessenen und möglichst spezifischsten Relationstypen ausgewählt werden.

Es kann aber auch vorkommen, dass die Erschaffung neuer, semantisch präziser Relationen die bestehende Relationierung nicht nur weiter differenziert, sondern generell in Frage stellt. In einem solchen Fall ist die Beibehaltung der allgemeinen Relationen sowie deren bloße Präzisierung nicht möglich, und es muss eine grundsätzlich neue Relationierung ausgewiesen werden. Durch eine solch grundlegende Neurelationierung unter Nichtbeachtung der alten Strukturen kann eine Einengung der gestalterischen Möglichkeiten durch gegebene Strukturen vermieden werden. Hierbei können punktuell Relationierungen entstehen, die den alten, überwundenen Strukturen oberflächlich zu entsprechen scheinen, welche sich aber vor allem in der formalen Definition ihrer logischen Eigenschaften von diesen unterscheiden.

Inwieweit die hierarchische Gestaltung des Relationeninventars sowie die schrittweise Entwicklung der spezifischen Relationen aus den bereits bestehenden Hauptrelationen eine parallele Modellierung alter und neuer relationaler Strukturen sowie deren gleichzeitige Präsentation für Indexierungs-

und Retrievalzwecke zulässt, ohne den Gebrauch der neuen spezifischen Relationen negativ zu beeinflussen und möglicherweise unbewusst einzuschränken, muss sehr genau untersucht werden. Die Verwendung zahlreicher neuer und hochgradig spezifischer Relationen und deren Einsatz zu einer umfangreichen, engmaschigen und semantisch präzisen Vernetzung des Begriffsraumes, in dem die *alten* Strukturen lediglich als krude Relikte unter der Oberfläche weiterexistieren, schmälert nicht deren Nutzen, solange deren Eigenschaften und Nachteile bewusst behandelt werden und der Gebrauch durch die Hinzunahme der neugeschaffenen Strukturen ergänzt wird.

So kann eine einfache, unspezifische Relationierung des Begriffsraums einen ersten allgemeinen Einstieg in den Gegenstandsbereich darstellen, an den sich eine differenziertere Erkundung anhand spezifischerer Beziehungen zwischen den Begriffen anschließt.

Da eine nachträgliche Integration und optionale Auswahl der ursprünglichen Relationierung durch die Kombination oder Vermischung alter Vokabularien und neuer Ontologien unter bestimmten technischen Voraussetzungen leicht möglich ist, besteht diese Darstellungsoption, beispielsweise für Illustrationszwecke oder Vergleiche immer noch.

Langfristig sollte aber eine vollständige Relationierung aller Bereich mit den neuen Relationen angestrebt werden, um somit eine umfassende und einheitliche Relationierung zu gewährleisten. Die so erschaffenen neuen Strukturen bilden als sorgfältig ausgearbeitete Ontologien die Grundlage für zahlreiche Anwendungsmöglichkeiten und Nutzungsszenarien. In welcher Form auf diese solchermaßen weiterentwickelten Dokumentationssprachen, gleich begrifflichen und strukturellen Fundamente aufgebaut wird, und welche zusätzlichen Überlegungen, Komponenten und Produkte für ein funktionierendes Gesamtkonzept benötigt werden oder entwickelt werden können, soll im sechsten Kapitel in einem Ausblick erörtert werden.

7 SCHNITTSTELLEN UND PRODUKTE

Die im vorangegangenen Kapitel dargestellten Möglichkeiten stellen theoretische Überlegungen zu neuen Gestaltungsmöglichkeiten von Dokumentationssprachen sowie der Strukturierung von Erschließungsvokabularen dar. Die dargestellte Relationierung ist als Kernbereich und Funktionsgrundlage der Erschließung zu verstehen, auf den zahlreiche Komponenten aufsetzen können. Einige dieser notwendigen oder möglichen Komponenten sollen im Folgenden angesprochen werden, um somit ein Bewusstsein dafür zu schaffen, welche weiteren Überlegungen sich an eine solche Arbeit anschließen und diese ergänzen können.

7.1 Schnittstellen für Visualisierungs- und Navigationskomponenten

Eine Dokumentationssprache ist für das Funktionieren eines Information Retrieval Systems von herausragender Bedeutung und spielt eine zentrale Rolle. Es wurde bereits darauf hingewiesen, dass eine isolierte Betrachtung der Dokumentationssprache nicht sinnvoll und problematisch ist. Die Gestaltung der Navigationskomponenten sowie die gesamte Visualisierung als Teil der Benutzeroberfläche müssen an die Erfordernisse des Systems angepasst werden, damit die Funktionalitäten, die, aufsetzend auf den begrifflichen Strukturen der Dokumentationssprache, zur Verfügung stehen, dem Nutzer in nachvollziehbarer und leichter handhabbarer Weise angeboten werden können. Dabei geht es nicht darum, die kompletten Strukturen in ihrer Gesamtheit mit all ihren Details zu zeigen, sondern einen selektiven und damit in seiner Komplexität überschaubaren Ausschnitt oder eine bestimmte Perspektive anbieten zu können. Zu Fragen der Visualisierung sei hier auf die Arbeiten von Sieber⁷⁶ und Trunk⁷⁷ verwiesen.

Die Verwendung eines weit verbreiteten Standards wie RDF/OWL erleichtert die Erstellung geeigneter Produkte und Komponenten. Auch die Integration und gemeinsame Nutzung verschiedener Terminologien oder Dokumentationssprachen, die sich entsprechend kodierter Strukturen bedienen, wird durch die Verwendung eines solchen Standards erleichtert. Gerade in Hinblick auf die immer wichtiger werdende Frage der semantischen Interoperabilität und der dadurch ermöglichten parallelen Nutzung mehrerer Dokumentationssprachen sind die Fragen einer differenzierten und formal

⁷⁶ Sieber, Wolfram. Thesaurus-Arbeit versus Informationsvisualisierung. Analyse und Evaluation am Maßstab der Usability. Saarbrücken: VDM, 2007.

⁷⁷ Trunk, Daniela. Semantische Netze in Informationssystemen. Verbesserung der Suche durch Interaktion und Visualisierung. 2005. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft Band 51)

streng ausgewiesenen Relationierung sowie deren entsprechende Kodierung und mögliche Nutzung von entscheidender Bedeutung.

7.2 Darstellung der Retrievalmöglichkeiten für den Benutzer

Ein wesentlicher Faktor bei der Gestaltung eines effizienten und leistungsstarken Information Retrieval Systems ist die angemessene und funktionale Gestaltung einer Schnittstelle zwischen System und Benutzer, die es jenem ermöglicht, das gesamte Potenzial der angebotenen Funktionalitäten von Dokumentationssprachen mit all ihren Leistungsmerkmalen für eine erfolgreiche Recherche zu nutzen. Gerade bei der Integration und Nutzung sehr komplexer und umfangreicher begrifflicher Strukturen sowie anspruchsvoller Retrievalfunktionalitäten ist eine genaue Abstimmung der technischen Realisation sowie der software-ergonomischen Gestaltung der Benutzeroberfläche auf die begrifflichen Strukturen des verwendeten Vokabulars besonders wichtig. Zur übersichtlichen und benutzerfreundlichen Darstellung eines umfangreichen Relationeninventars sollten die hierarchischen Strukturen des Inventars genutzt werden. Für die Darstellung aller zur Relationierung eines bestimmten Begriffes tatsächlich verwendeten Relationstypen könnte man dieses durch ein speziell auf den jeweils ausgewählten Begriff abgestimmtes kontextsensitives Auswahlmenü erreichen, welche jeweils nur die benötigte Auswahl an Relationstypen und nicht das gesamte Inventar zur Auswahl anbietet

7.3 Weiterführende Überlegungen / Ausblick

Auch für weiterführende Überlegungen zu der Gestaltung von Dokumentationssprachen im Hinblick auf deren Verwendung in Systemen oder Anwendungsszenarien welche eine größtmögliche semantische Interoperabilität anstreben, bieten die geleisteten Überlegungen und Untersuchungen wichtige Anknüpfungspunkte. Die Differenzierung und Austypisierung ist eine wichtige Voraussetzung für den Versuch, verschiedene Dokumentationssprachen auch aus unterschiedlichen Sprachräumen miteinander in Beziehung zu setzen und zu verbinden. Die Frage nach allgemeingültigen, möglicherweise sogar sprachunabhängigen aber dennoch hinreichend spezifischen Relationeninventaren zur Gestaltung verschiedener Dokumentationssprachen auch im Hinblick auf Fragen semantischer Interoperabilität multilinguale Anwendungsmöglichkeiten, können an die angestellten Überlegungen und gewonnenen Erkenntnisse anknüpfen.

8 SCHLUSS

Die hier angestellten Überlegungen und Vorschläge zur Verfeinerung und Differenzierung semantischer Relationen in kontrollierten Vokabularen von Dokumentationssprachen zeigen, wie spezifische semantische Relationen zu Zwecken der semantischen Modellierung entwickelt und verwendet werden können. Die Entwicklung dieser semantischen Relationen aus dem bestehenden Vokabular einer verwendeten Dokumentationssprache ermöglicht die Weiterverwendung der etablierten Erschließungsinstrumente sowie der erschlossenen Bestände. Die stufenweise semantische Anreicherung solcher Dokumentationssprachen sind die Voraussetzung für deren sinnvolle Weiterverwendung im Bereich technisch hoch entwickelter Information Retrieval Systeme in elektronischen Umgebungen. Die Verwendung einer induktiven Methode, die das Inventar spezifischer und präzise definierter Relationstypen durch eine stufenweise Abstraktion von Einzelrelationen innerhalb eines Gegenstandsbereiches entwickelt und durch weitergehende Verallgemeinerung für den Gebrauch in verschiedenen Themengebieten aufbereitet, bietet die Chance zur Entwicklung eines Inventars dessen Relationen sich gleichzeitig durch eine hinreichende Allgemeinheit sowie eine größtmögliche Spezifität auszeichnen und so eine aussagekräftige Modellierung zahlreicher Gegenstandsbereiche ermöglichen.

Es zeigt sich, dass die Weiterentwicklung und semantische Anreicherung bestehender Dokumentationssprachen zu informationell gehaltvolleren und somit aussagekräftigeren Strukturen nicht nur notwendig sondern auch möglich sind. Die genaue Analyse eines bestehenden Begriffsraumes zur Entwicklung und Gestaltung semantisch differenzierter Relationstypen, sowie deren präzise Definition ergeben hier ein spezifisches und auf den Gegenstandsbereich abgestimmtes und in diesem verankertes Relationeninventar. Die dadurch ermöglichte Neugestaltung bereits etablierter und genutzter Dokumentationssprachen auf struktureller Ebene kann die Einsatzmöglichkeiten dieser Sprachen und Vokabularen erweitern. Diese können auf vielfältige Weise weiter genutzt werden. So werden semantisch differenziert relationierte Sprachen zur Funktionsgrundlage eines leistungsstarken Information Retrieval und ermöglichen somit eine genauere, zielgerichtete Orientierung und erfolgreichere Suche im Begriffsraum.

Die hier angestellten Überlegungen verdeutlichen die Notwendigkeit weiterer Anstrengungen bei der Entwicklung von Richtlinien und Standards für die Entwicklung differenzierter Relationen und Relationeninventare.

Auch für weitergehende Überlegungen bezüglich Fragen multilingualer Anwendungen oder entlokalisierter sprachunabhängiger Begriffsstrukturen sind Methoden zur Entwicklung und präzisen Austypisierung semantischer Relationen innerhalb solcher Systeme eine wichtige Voraussetzung und bieten ein großes Potenzial für weitere Forschung.

LITERATURVERZEICHNIS

- Aitchison, Jean; Alan Gilchrist, David Bawden. Thesaurus construction and use: a practical manual. London: Aslib 1997.
- American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSS Subject Analysis Committee. (1997)
<http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/finalreport.htm> [Letzter Zugriff 17.02.2008]
- Bean, Carol A. Rebecca Green (Hrsg.): Relationships in the Organisation of Knowledge. Dordrecht: Kluwer Academic Publishers. 2001 (Information Science and Knowledge Management 2)
- Beghtol, C.: "The facet concept as a universal principle of subdivision" In: Knowledge organization, information systems and other essays: Professor A. Neelamegham Festschrift. K.S. Raghavan K.N. Prasad. (Hrsg.) New Delhi: Ess Ess Publications 2006. S.41-52.
- Bertram, Jutta. Einführung in die inhaltliche Erschließung. Grundlagen, Methoden, Instrumente. Würzburg: Ergon-Verl. 2005.
- Budin, Gerhard. Wissensorganisation und Terminologie. Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. Tübingen: Narr, 1996. (Forum für Fachsprachenforschung 28).
- Clarke, Stella G. Dextre. „Thesaural Relationships.“ In: Bean, Carol A. Rebecca Green (Hrsg.): Relationships in the Organisation of Knowledge. Dordrecht: Kluwer Academic Publishers. 2001 (Information Science and Knowledge Management 2). S.37-52
- Dee, Michel, Taxonomy of subject relationships, Appendix B zu: American Library Association, Subcommittee on Subject Relationships / Reference Structure (Hrsg.) Final Report to the ALCTS/CSS Subject Analysis Committee. (1997)
<http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/msrscu2.pdf> [Letzter Zugriff 17.02.2008]
- Efthimiadis, E. N. "Query expansion" In: Annual Review of Information Science and Technology 31 (1996) S. 121-187
- Efthimiadis, E. N. "Interactive query expansion: A user-based evaluation in a relevance feedback environment" Journal of the American Society for Information Science 51, (2000) S. 989-1003
- Fischer, Dietrich H. "From Thesauri towards ontologies?" In: Advances of Knowledge Organization, vol. 6 (1998) S. 18-30
- Fischer, Karl-Heinz. „Der BERLIN-THESAURUS als Beispiel für einen komplexen Regionalthesaurus“ In: Nachrichten für Dokumentation 39 (1988) S. 245-252
- Foskett, A. C. *The subject approach to Information*. 5th edition, London: Library Association Publishing, 1996
- Gödert, Winfried. „Externalisierung von Wissen: Eine informationstheoretische Betrachtung aus konstruktivistischer Sicht“ In: Meder, Norbert; Peter Jaenecke, Winfried Gödert (Hrsg.): Konstruktion und Retrieval von Wissen. 3. Tagung der Deutschen ISKO Sektion einschließlich der Vorträge des Workshops „Thesauri als terminologische Lexika.“ Weil-

-
- burg, 27.-29. 10. 1993. Frankfurt a.M.: Indeks-Verlag 1995. (Fortschritte in der Wissensorganisation 3)
- Gödert, Winfried. „Verbale Inhaltserschließung : Ein Übersichtsartikel als kommentierter Literaturbericht“ In: Mitteilungen für Bibliotheken des Landes Nordrhein-Westfalen, 41 (1991) 1, S. 1-21
- Gödert, Winfried; Klaus Lepsky. Semantische Umfeldsuche im Information Retrieval in Online-Katalogen. Köln: Fachhochschule Köln, Fakultät für Informations- und Kommunikationswissenschaften, Institut für Informationswissenschaft 1998. (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft 7)
- Green, Rebecca. „Syntagmatic relationships in index languages: a reassessment“. In: Library quarterly. 65 (1995) (4). S. 365-385.
- Green, R., C.A. Bean: “Aligning systems of relationships”. In: Knowledge organization, information systems and other essays: Professor A. Neelamegha Festschrift. K.S. Raghavan and K.N. Prasad. (Hrsg.) New Delhi: Ess Ess Publications 2006. S.111-128.
- Hubrich, Jessica. Input und Output der Schlagwortnormdatei (SWD). Aufwand zur Sicherstellung der Qualität und Möglichkeiten des Nutzens im OPAC. Master Thesis. (2005) (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft Band 49)
- Hutchins, W. John. Languages of indexing and classification. A linguistic study of structures and functions. Stevenage: Peregrinus 1975
- Kuhlen, Rainer; Thomas Seeger; Dietmar Strauch (Hrsg.) Grundlagen der praktischen Information und Dokumentation. 5., völlig neu gefasste Ausgabe. München: Saur, 2004.
- Kunz, Martin. „Zur Weiterentwicklung der SWD“ In: Bibliotheksmanagement – Kulturmanagement. Vorträge und Berichte. 24. Österreichischer Bibliothekartag Congress Innsbruck , 3-7.9. 1996 (Biblos-Schriften Band 168) S.177-196
- Larivière, L. „Vers un produit unifié en terminologie et en documentation: le thésaurus terminologique.“ In: Méta, 34 (1989) S. 457-467
- Maistra, Ranjita. Semi-automatic method of preparing thesaurus for a specific subject-field. Part B of the FID/CR Report No 18 of the Committee on Classification Research (FID/CR) International Federation for Documentation (FID). Bangalore: Documentation Research and Training Centre 1978
- Mazzocchi, Fulvio; Melissa Tiberi, Barbara De Santis, Paolo Plini. “Relational semantics in thesauri: some remarks at theoretical and practical levels”. In: Knowledge Organization Vol. 34 (2007) 4 S. 197-214
- Mustafa El-Hadi, Widad (Hrsg.). Structures and Relations in Knowledge Organisation .25.-29. August 1998 Lille, France.Organized by UFR IDIST University Charles de Gaulle Lille III and the ISKO General Sekretariat. Würzburg: Ergon-Verlag 1998. (Advances in Knowledge Organisation 6.
- Neelamegha, A; I. K. Ravichandra Rao. Non-hierarchical associative relationships: their types and computerization of links. In: Library science with a slant to documentation and information studies (1976) 13, S. 24-42
- Neelamegha, A; Ranjitra Maistra. Non-hierarchical associative Relationships among concepts: identification and typology. (Part A of the FID/CR Report No 18 of the Committee on Classification Research (FID/CR) International Federation for Documentation (FID) Bangalore: Documentation Research and Training Centre 1978

-
- Pellegrini, Tassilo; Andreas Blumauer Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen . In: Pellegrini, Tassilo; Andreas Blumauer (Hrsg.) *Semantic Web. Wege zur vernetzten Wissensgesellschaft*. Berlin [u.a.] Springer, 2006 S. 9-24.
- Pellegrini, Tassilo; Andreas Blumauer (Hrsg.) *Semantic Web. Wege zur vernetzten Wissensgesellschaft*. Berlin [u.a.] Springer, 2006
- Peters, Isabella; Katrin Weller. "Paradigmatic and syntagmatic relations in knowledge organization Systems." In: *Information Wissenschaft und Praxis*. 59 (2008) 2 S.100-107
- Reimer, Ulrich. *Einführung in die Wissensrepräsentation*. Stuttgart: Teubner, 1991 (Leitfäden der angewandten Informatik)
- Reimer, Ulrich. „Wissensbasierte Verfahren der Organisation and Vermittlung von Information“. In: Kuhlen, Rainer; Thomas Seeger; Dietmar Strauch (Hrsg.) *Grundlagen der praktischen Information und Dokumentation*. 5., völlig neu gefasste Ausgabe. Bd. 1 *Handbuch zur Einführung in die Informationswissenschaft und –praxis* München: Saur, 2004. S. 155-166
- Sager, J. C. *A practical course in terminology processing*. Amsterdam: Benjamins. 1990
- Schmid, Ute; Martin Christof Kindsmüller. *Kognitive Modellierung. Eine Einführung in die logischen und algorithmischen Grundlagen*. Heidelberg [u. a.]: Spektrum Akademischer Verlag, 1996
- Schöndorf, Peter. „Nicht-konventionelle Thesaurusrelationen als Orientierungshilfe für Indexierung und Recherche – Analyse ausgewählter Beispiele.“ In: *Nachrichten für Dokumentation* 39 (1988) S. 231-244
- Sieber, Wolfram. *Thesaurus-Arbeit versus Informationsvisualisierung. Analyse und Evaluation am Maßstab der Usability*. Saarbrücken: VDM, 2007.
- Soergel, Dagobert. *Klassifikationssysteme und Thesauri. Eine Anleitung zur Herstellung von Klassifikationssystemen und Thesauri im Bereich der Dokumentation*. Frankfurt am Main: Deutsche Gesellschaft für Dokumentation 1969.
- Sowa, John F. *Knowledge representation. logical, philosophical, and computational Foundations*. London [u. a.]: Brooks/Cole. 2000.
- Sowa, John F. *Principles of semantic networks. Explorations in the representation of knowledge*. San Mateo (CA): Morgan Kaufmann Publishers; 1991
- Swenonius, Elaine. *The intellectual foundation of information organization*. Cambridge, MA: The MIT Press. (2000)
- Tudhope, Douglas; Harith Alani; Christopher Jones. "Augmenting thesaural relationships: Possibilities for Retrieval". In: *Journal of Digital Information*, 1 (8) (2001) unter <http://jodi.tamu.edu/Articles/v01/i08/Tudhope/> [Letzter Zugriff: 17.02.2008]
- Trunk, Daniela. *Semantische Netze in Informationssystemen. Verbesserung der suche durch Interaktion und Visualisierung*. 2005. *Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft* Band 51.
- Wang, Yih-Chen; James Vandendorpe, Matha Evens. "Relational Thesauri in information retrieval." In: *Journal of the American Society of Information Science*. 36 (1) (1985) S. 15-27.
- Wersig, Gernot. *Thesaurus-Leitfaden. Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis*. München: Saur, 1985 (DGD-Schriftenreihe 8)

Internetquellen

CrissCross Projekt:

<http://www.fbi.fh-koeln.de/institut/projekte/CrissCross/index.html>

[letzter Zugriff: 06.03.2008]

MACS Projekt:

<https://macs.vub.ac.be/pub/>

[letzter Zugriff: 06.03.2008]

Protégé Protégé: The Protégé Ontology Editor and Knowledge Acquisition System

<http://protege.stanford.edu/>

[letzter Zugriff: 06.03.2008]

Referenzwerke

Manfred Brauneck; Gerhard Schneilin. Theater Lexikon. Begriffe und Epochen/Bühnen und Ensembles. Reinbek bei Hamburg: Rowohlt, 1986

Jochanan Ch. Trilse-Finkelstein, Klaus Hammer Lexikon Theater International.. Berlin: Henschel, 1995

Rischbieter, Henning (Hrsg.). Theater-Lexikon. Zürich [u. a.] Orell Füssli, 1983

Unveröffentlicht

Semantische Anreicherung der Schlagwortnormdatei. Winfried Gödert [u. a.] Ergebnisbericht Köln: Fachhochschule Köln. Institut für Informationswissenschaft 2004

ANHANG

- Differenzierung der bestehenden UB-Relationen für den Deskriptor 207 „Theater“
- Schaubilder zur Visualisierung von Suchanfragen
- Dateiverzeichnis auf dem Datenträger

Differenzierung der bestehenden UB-Relationen für den Deskriptor 207 „Theater“:

UB generisch / Form (mögliche Überschneidung / Unschärfe mit UB generisch / Funktion)	<ul style="list-style-type: none"> - Agitproptheater - Alternatives Theater - Antitheater - Commedia dell' arte - Mundarttheater - Revuetheater - Sprechtheater - Théâtre de silence - Unterhaltungstheater - Volkstheater - Studio –Theater ?
UB generisch / Funktion	<ul style="list-style-type: none"> - Unterhaltungstheater - Politisches Theater - Fronttheater ? - Exiltheater ?
UB generisch / Beteiligte	<ul style="list-style-type: none"> - Arbeitertheater - Frauentheater - Jugendtheater - Kindertheater - Schulspiel - Studententheater - Exiltheater
UB generisch / Professionalität	<ul style="list-style-type: none"> - Berufstheater - Laienspiel
UB generisch / Institution	<ul style="list-style-type: none"> - Freies Theater ? - Hoftheater - Landesbühne - Stadttheater - Staatstheater
UB generisch / Publikum	<ul style="list-style-type: none"> - Arbeitertheater - Hoftheater - Jugendtheater - Kindertheater - Schulspiel
UB generisch / Methode	<ul style="list-style-type: none"> - Papiertheater - Playbacktheater - Lesebühne - Figurentheater - Experimentelles Theater - Studio-Theater

UB generisch / Thematik	<ul style="list-style-type: none">- Erotisches Theater- Metatheater- Politisches Theater
UB generisch / Lokaler Kontext	<ul style="list-style-type: none">- Boulevardtheater ?- Freilichttheater- Fronttheater- Hoftheater- Kammertheater- Kellerbühne- Provinztheater ?- Straßentheater- Zimmertheater

Schaubilder zur Visualisierung von Suchanfragen:

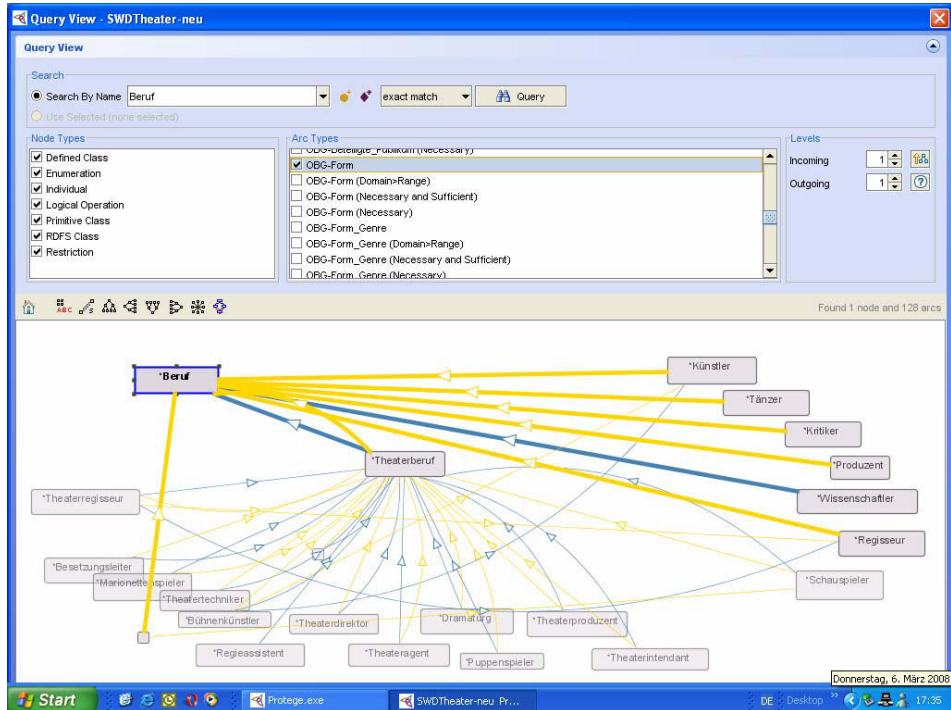


Abbildung 10: Selektive Relationierung des Terms "Beruf"

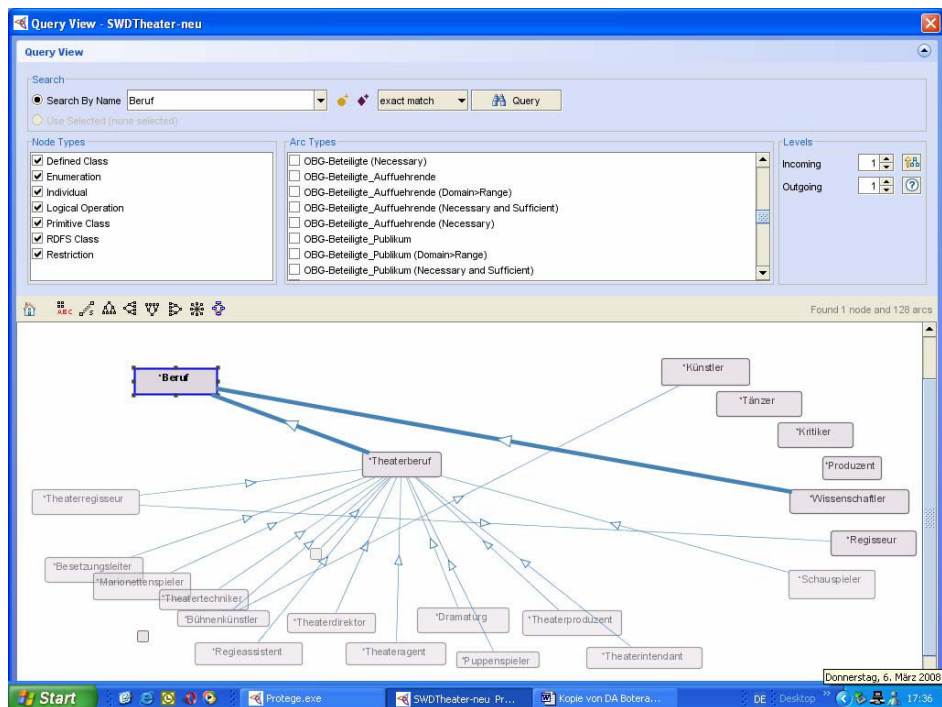


Abbildung 11: Selektive Relationierung des Terms "Beruf"

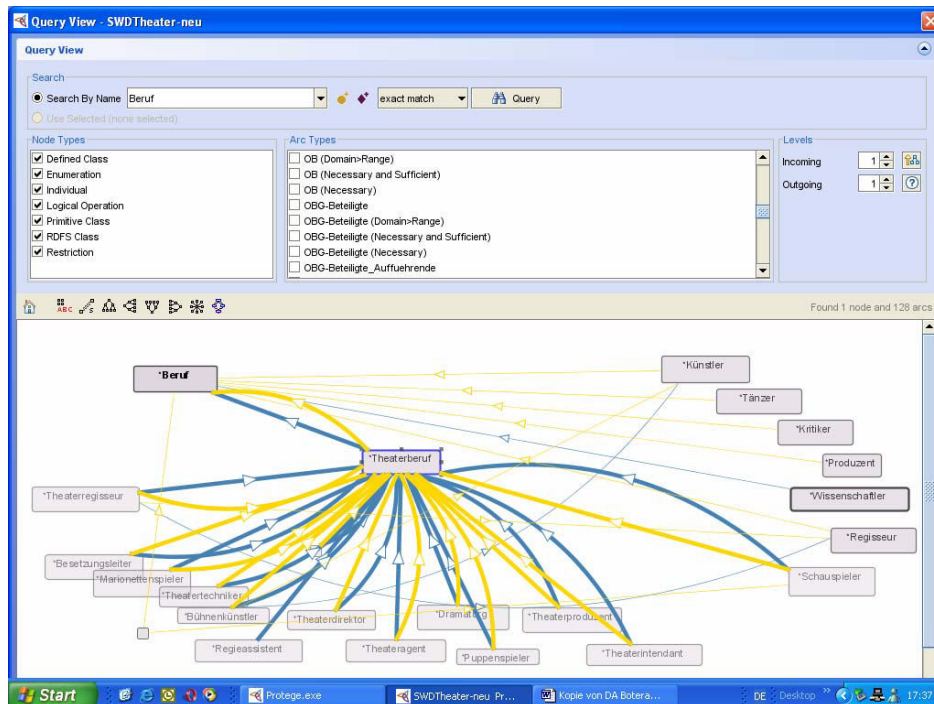


Abbildung 12: Hierarchieleiter zum Term "Beruf" mit Relationstypen unterschiedlicher Spezifität

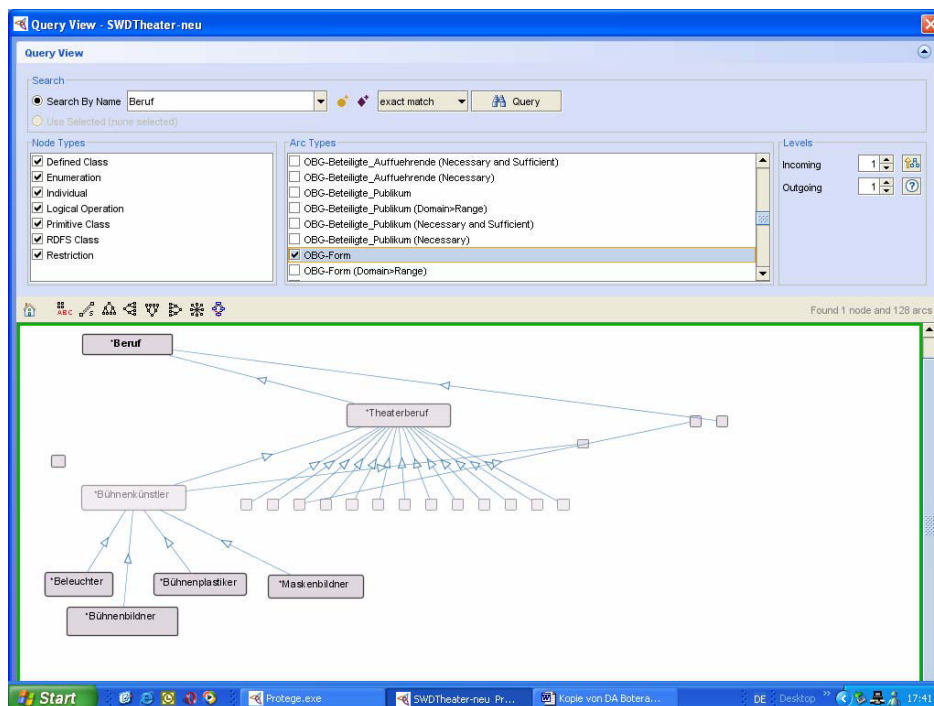


Abbildung 13: Hierarchieleiter zum Term "Beruf"

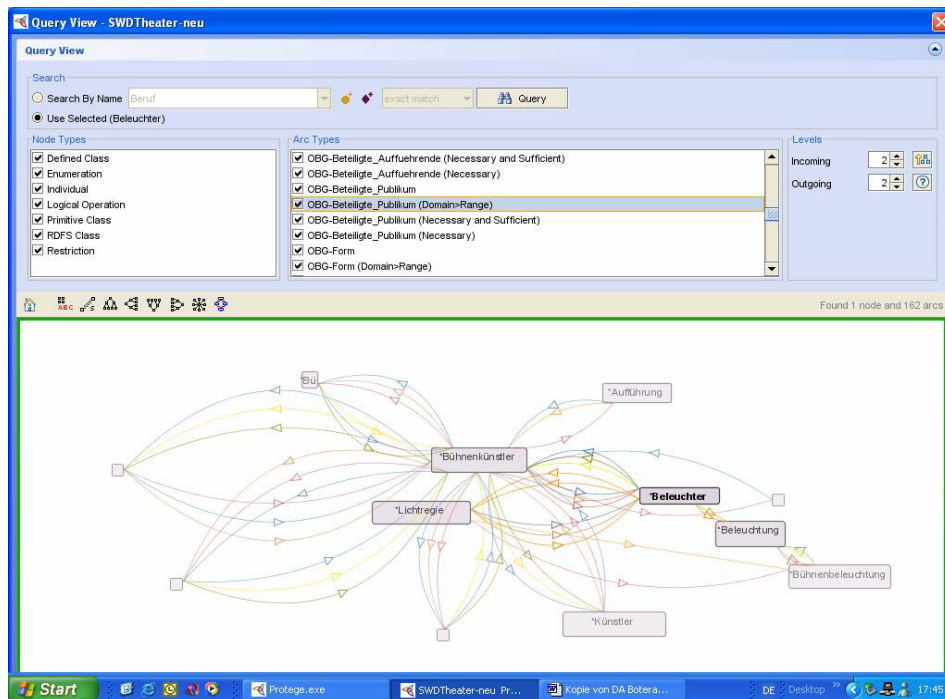


Abbildung 14: Semantisches Umfeld des Terms "Beleuchter"

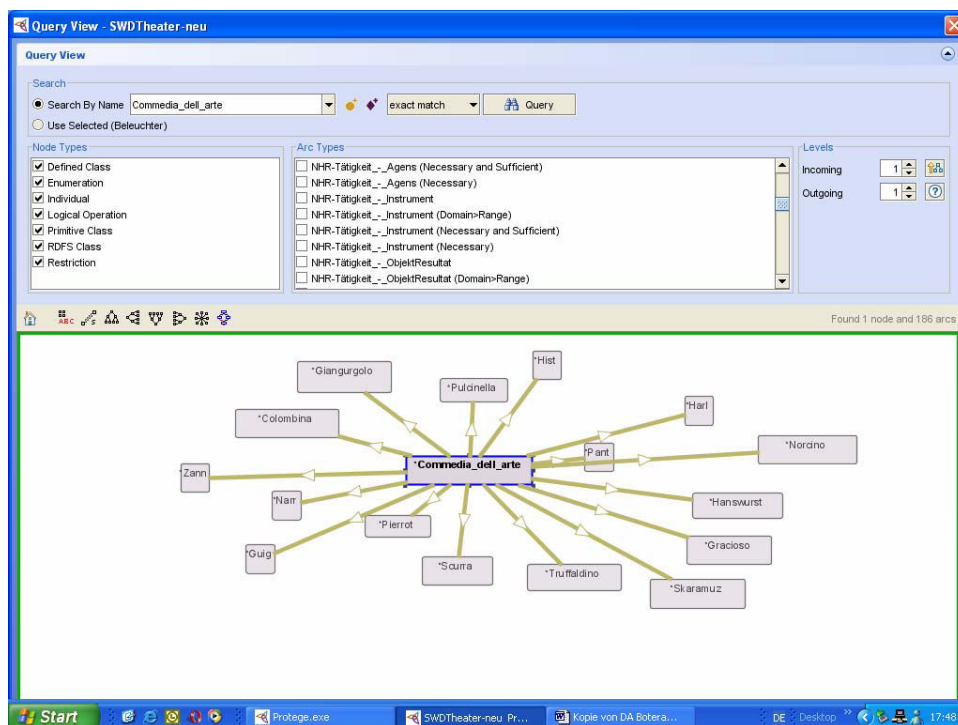


Abbildung 15: Relation "Beteiligte" zum Term "Commedia dell'arte"

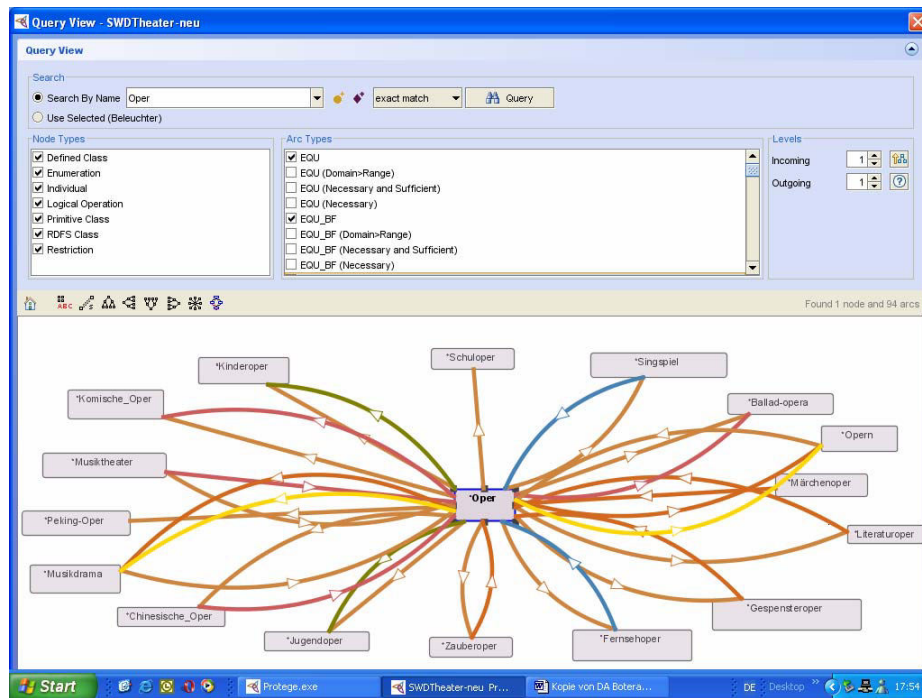


Abbildung 16: Semantisches Umfeld des Terms "Oper"

Dateiverzeichnis des Datenträgers CD1:

Kürzel	Dateiname	Format	Beschreibung
DA1	Diplomarbeit Boteram [...].pdf	pdf	Textdokument der Diplomarbeit
ABST	Abstract.rtf	rtf	Abstract der Diplomarbeit
SWD1	SWDExport.txt	txt	Auszug aus der SWD (Sachgruppe 15.1) über Notationstrunkierung im MAB-Format
SWD2	SWDExport.mth	mth	Auszug aus der SWD (Sachgruppe 15.1) im mth.Format (strukturell identisch mit SWD1)
MACS1	MACSTheater	pdf	Deskriptorenliste SWD Bereich „Theater“ des MACS Projekts
PRO1	SWDTheater-alt	owl	Übertrag der Deskriptorenliste inklusive der in der SWD ausgewiesenen Relationen in OWL (strukturell identisch mit PRO2)
PRO2	SWDTheater-alt	pprj	Übertrag der Deskriptorenliste inklusive der in der SWD ausgewiesenen Relationen in das Protégé-Projekt „SWDTheater-alt“ (strukturell identisch mit PRO1)
PRO3	SWDTheater-neu	owl	Neurelationierung von SWDTheater-alt
PRO4	SWDTheater-neu	pprj	Neurelationierung von SWDTheater-alt
REL1	Relationeninventar.mth	mth	Relationeninventar für SWDTheater-neu.owl/pprj
REL2	TaxonomyDee	mth	Relationeninventar Michel Dee